Criminal Justice Statistical Analysis Center

July 2011

State of West Virginia
Department of Military Affairs & Public Safety
Division of Justice and Community Services
Office of Research and Strategic Planning

Assessing the Validity of Hate Crime Reporting: An Analysis of NIBRS Data

Stephen M. Haas, Ph.D., ORSP Director
James J. Nolan, Ph.D., West Virginia University
Erica Turley, B.S., ORSP Research Analyst
Jake Stump, B.S., West Virginia University

The Uniform Crime Reporting (UCR) Program was developed over 80 years ago to meet the need for reliable crime statistics for the nation. Today, nearly 17,000 law enforcement agencies across the United States participate in this voluntary program. UCR, and the modernized National Incident-Based Reporting System (NIBRS), are recognized as the primary source of information about crimes reported to the police. While the UCR Program is critical to understanding crime, there are known limitations to these data such as underreporting and misclassification.

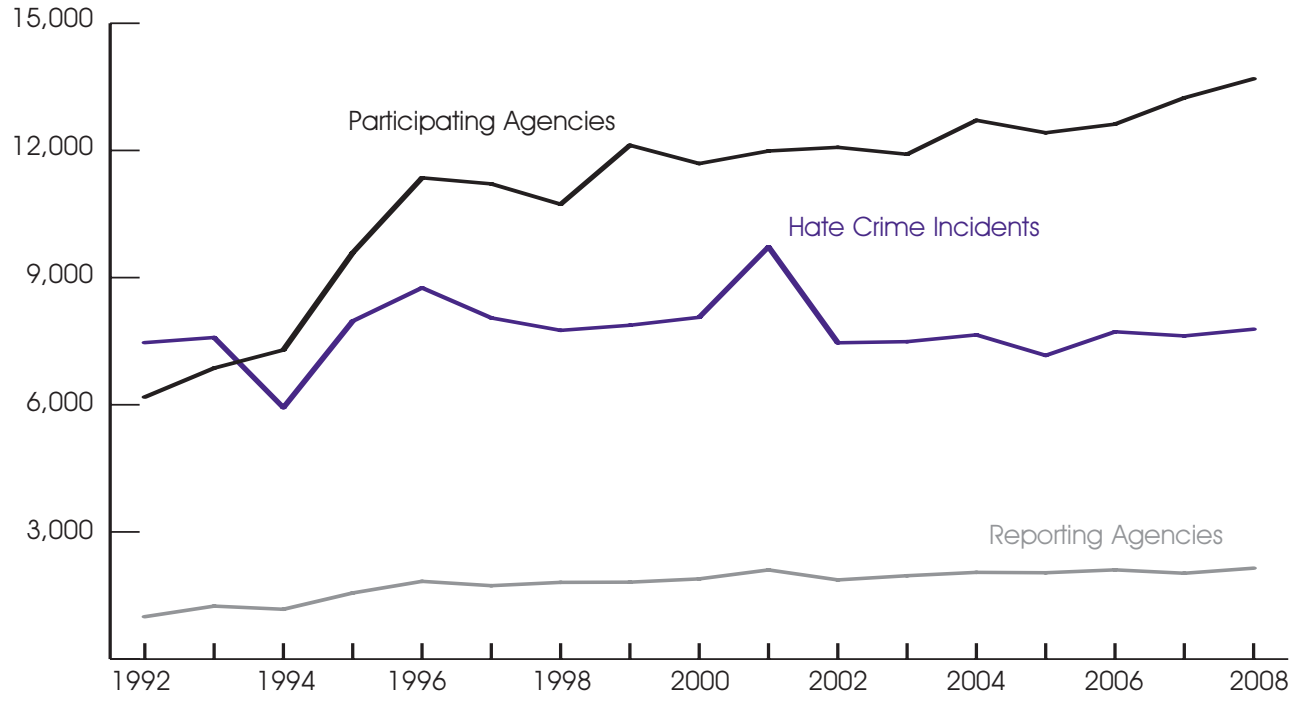
As with any large scale data collection system, errors are inevitable and occur for a variety of reasons. While it is unlikely that all error will be eliminated, it is important to understand and measure it. Classification error occurs when the facts of the crime are recorded by the police, but the crime type is identified incorrectly. These errors can occur for many reasons including inaccurate interpretation of UCR definitions, reliance on criminal rather than statistical definitions, record automation issues, and even purposive actions in an attempt to downgrade crime. Classification error is particularly important since it can ultimately impact the statistical accuracy of reported crime statistics.

The purpose of the current study is to examine the misclassification of crimes as they relate to hate, that is, the degree to which classification error impacts the statistical accuracy of reported hate crimes. Such error can vary by crime type and result in both the undercounting and overcounting of crimes in official statistics. To focus this study on hate crimes is noteworthy because, by their very nature, hate crimes create a unique set of classification issues. Inherently, the intention of people involved and/or their motivation for committing a crime must be taken into account by officers when determining whether a particular incident constitutes a hate crime. For this reason, and others to be discussed later in this report, it is often speculated that many hate crimes are not accurately recorded in official records. Through a systematic review of official records, this study seeks to examine the degree to which classification error impacts the statistical accuracy of hate crime, as reported in official law enforcement statistics.

Utilizing a methodology previously developed by the authors (Nolan, Haas, and Napier, 2011; Nolan, Haas, Lester, Kirby, and Jira, 2006), this study assesses the amount of classification error in hate crime reporting in West Virginia. The researchers randomly selected cases, which were included in the state's statistical data files, from designated offense categories for a detailed review of the officer's written narrative of the incident. Though this approach has been applied to examine error across general crime types, no study to date has systematically focused on a crime category as widely believed to be underreported as hate crimes. While the previous study examined classification error across general crime types, the current study focuses specifically on identifying sources of error (i.e., over- and undercounts) contained in hate crime statistics.

Additionally, this study builds on the quantitative method described above by further capturing the perspectives of frontline officers. Qualitative information from a focus group is used to gain insight into the thought processes officers adhere to when deciding whether a specific incident constitutes a hate crime. Equipped with narratives of cases believed to contain errors, the researchers

Figure 1. Participation and Reporting Trends in FBI National Hate Crime Program, 1992-2008



Source: FBI Hate Crime Statistics

use a focus group approach to explore the various definitional and interpretation issues that are believed to result in classification error in these cases. Thus, it is anticipated that this study will not only yield an estimate of the error contained in officially reported hate crime statistics, but shed light on the inherent difficulties officers face in interpreting these incidents. In the end, it is the hope of the authors that this study will yield useful information for training officers on the reporting of hate crimes, get us closer to understanding the true magnitude of these crimes, and serve as a precursor for adjusting crime statistics to better estimate the actual number of hate crimes in the population.

The National Hate Crime Data Collection Program

The FBI established the national hate crime data collection program in response to the Hate Crime Statistics Act (HCSA) of 1990. The HCSA required the attorney general to establish guidelines and collect data "about crimes that manifest evidence of prejudice based on race, religion, sexual orientation, or ethnicity, including, where appropriate, the crimes of murder, non-negligent manslaughter; forcible rape; aggravated assault, simple assault, intimidation; arson; and destruction, damage, or vandalism of property (Public Law 101–275)." The attorney general appointed the director of the FBI with the responsibility for developing a national data collection program for hate crimes. Working with other law enforcement officials, criminologists, nongovernmental organizations (NGOs), and a myriad of other supporters, the FBI developed the current national hate crime program as an adjunct to the existing—and well-established—Uniform Crime Reporting (UCR) program (Federal Bureau of Investigation [FBI], 1997).

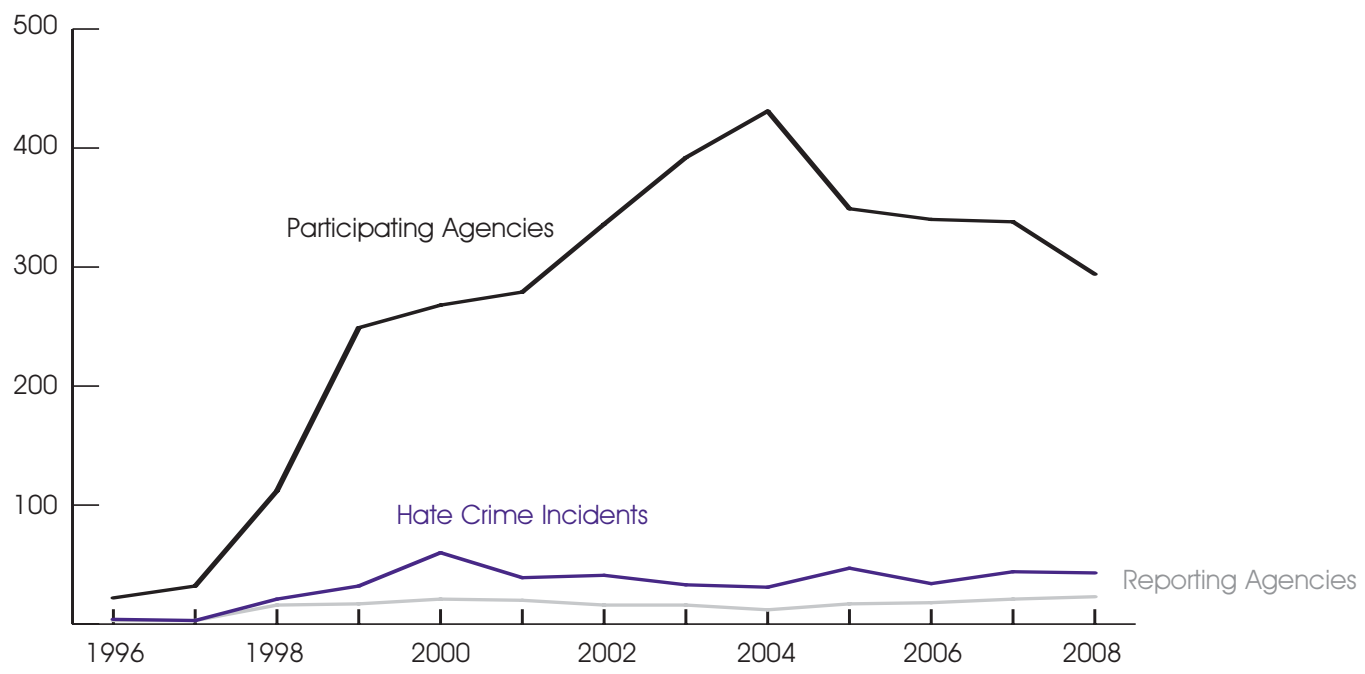
The voluntary participation of police agencies in the national data collection grew rapidly after 1992, when the first full publication of hate crime data was published by the FBI (see Figure 1). In fact, in the 10 years following the first hate crime publication, law enforcement participation doubled, from roughly 6,000 police agencies to 12,000. In 2008, the most recent year with published hate crime data, nearly 14,000 police agencies participated in the program. Similarly, participation among police agencies in West Virginia (the location of this study) grew at a rapid pace annually (see Figures 2 and 3). Only 22 agencies in West Virginia participated in the program in 1996. By 2004, the number peaked at 431. Since 2004, West Virginia has seen

Figure 2. Hate Crime Reporting by West Virginia Police Agencies, 1996-2008

Year	Participating Agencies	Population Covered by Agencies	Agencies Reporting Hate Crimes	# of Hate Crime Incidents
1996	22	179,467	3	4
1997	32	367,499	3	3
1998	112	708,363	16	21
1999	249	1,466,686	17	32
2000	268	1,669,624	21	60
2001	279	1,756,724	20	39
2002	336	1,758,307	16	41
2003	392	1,712,606	16	33
2004	431	1,815,354	12	31
2005	349	1,679,815	17	47
2006	340	1,707,846	18	34
2007	338	1,686,872	21	44
2008	294	1,636,800	23	43

Source: FBI Hate Crime Statistics

Figure 3. Hate Crime Reporting by West Virginia Police Agencies, 1996-2008



Source: FBI Hate Crime Statistics

annual declines in the number of participating agencies, yet the latest figure in 2008 shows nearly 300 agencies participated, covering a population of 1.6 million.

It is important to understand that "participation" in the program does not mean that all of these law enforcement agencies investigated and reported hate crimes to the FBI. In fact, most of them did not. In order to be a participant in the FBI's program, a police agency must complete a hate crime incident report or submit a form signed by the chief of police that states no hate crimes have occurred that year in its jurisdiction. After 1996, the number of "reporting agencies" (i.e., those that actually experienced a hate crime and reported it to the FBI) has remained relatively stable at around 2,000. It is interesting to observe that the number of hate crime incidents has also remained relatively stable since the late-1990s, with the exception of a spike in 2001 (see Figure 1). This one-year spike in hate crime data reflects bias-motivated offenses that occurred following September 11, 2001, many of which were directed at Middle Eastern victims. Similarly, in West Virginia both the number of agencies that report hate crime and the number of hate crimes themselves have remained relatively stable since 1998.

Hate Crimes in West Virginia: Two Case Examples

Two celebrated bias-related cases in West Virginia may help illustrate the concept of hate crime in a localized setting. Within one decade, two homegrown cases involving murder, torture and rape generated headlines across the globe. These specific incidents also illustrated how local law enforcement struggles to correctly classify hate crimes.

These two examples (highlighted in the boxes below and at the top of page 5) showcase the complications surrounding hate crimes. It is anticipated that law enforcement's handling of these publicized cases represents its overall interpretation of bias-motivated events. If officers fail to identify hate crimes in serious cases involving murder or torture, it is expected they would not detect bias in low-level offenses or incidents lacking news appeal.

Hate Crime Reporting: Issues of Definition and Interpretation

Despite the advancement of hate crime laws and definitions over the last several decades, distinguishing bias-motivated crimes from unbiased crimes remains a difficult practice for law enforcement officials and even experts in the field. Hate crime remains a relatively new term – it emerged in the 1990s largely due to the FBI, which defined it as a criminal offense motivated 'in whole or in part' by bias. While the FBI definition serves as the standard for identifying a bias-motivated offense, applying it to the real world of a police officer can be downright confusing and problematic. This isn't to say law enforcement is the only profession that struggles with the concept of hate crime. Various public actors – including attorneys, elected officials, journalists and even scholars who study hate crime – wrestle with accurately identifying this special type of crime.

Defining Hate Crime

The FBI defines a hate crime as a "criminal offense committed against a person or property which is motivated, in whole or in part, by the offender's bias against race,

The Case of Arthur "J.R." Warren...

July 3, 2000. Arthur "J.R." Warren, a 26-year-old black gay man with learning disabilities, told his parents he was going to watch the Fourth of July fireworks in Grant Town, a rural municipality of 700 people. Instead, Warren ended up meeting three teenage acquaintances—David Parker and Jared Wilson, both 17, and Jason Shoemaker, 15. In an empty house, the four of them drank beer, smoked marijuana and huffed gasoline.

Later that evening, an argument ensued between Warren and Parker, reportedly over Warren revealing to other people that the two shared a sexual relationship. The argument turned physical when Parker and Wilson started punching Warren and kicking him with steel-toed boots. After the attack, the boys placed the bloodied victim into a car and drove. They stopped in a secluded area and dragged Warren, who was still conscious, into the middle of the road. Parker ran over Warren four times, killing him in an attempt to disguise the fatality as a hit-and-run.

The Case of Megan Williams...

September 2007. In rural Big Creek (Logan County), six local white residents kidnapped 20-year-old Megan Williams, a black woman who also had learning disabilities. She was held captive in a rundown trailer for a week and forced to eat animal feces, lick blood and drink from a toilet. Investigators say she was also sexually assaulted as her attackers hurled racial slurs at her. One of the accused, Bobby Brewster, was Williams' boyfriend. Williams survived the horror and escaped when police, acting on a tip from neighbors, arrived at the residence. As officers questioned Bobby Brewster's mother on the front porch, Williams stumbled out of the trailer and uttered "help me."

The victim's family members and prominent civil rights activists, including Al Sharpton, called for hate crime charges against all six defendants. Only one, Karen Burton, was charged and convicted of a hate crime. Although prosecutors racked up one hate crime conviction, police failed to initially report it as a bias crime, as it did not appear in the annual *FBI Hate Crime Statistics* publication.

In 2009, Williams recanted her story.

religion, disability, ethnic/national origin group, or sexual orientation group (FBI, 1997, p. 4)." In addition to this definition, states and localities have their own definitions and statutes regarding hate crime, which complicates the classification of hate crimes and will be explained in detail later in this report. For example, the WV law does not include disability or sexual orientation as protected classes but does provide protection against crimes based on political affiliation and gender. Since they are included in the legal definition in West Virginia they are also being captured in WV statistics. When submitted to the FBI, hate crimes in West Virginia involving gender or political affiliation are removed from the statistics. For the purposes of this study, both the FBI definition and the additional two classes included in West Virginia were considered during case review. In addition to definitional issues, various typologies exist for interpreting whether hate crimes occurred. The following section examines these types of issues that law enforcement officers often struggle with.

Typology of Hate Crime Reports. New words and phrases emerge all the time. From the social sciences we know that this process of creating new terms often takes the form of a two-stage process: *intension* and *extension* (Dewey, 1910|1997). First, a term is created and defined (intension) and then it is applied to real-life events (extension). The process continues until a shared understanding of the term is achieved.

This process of intension and extension relating to hate crime reporting by the police was examined by James Nolan and his colleagues in 2004 (Nolan, McDevitt, Cronin, and Farrell, 2004). This study of hate crime reporting by police, found that the FBI's definition of hate crime, particularly "motivated in whole or in part by bias" created ambiguity and often frustrated officers who wanted to report hate crimes accurately. Specifically, there were two types of crimes that gave officers the most trouble: 1) Response/Retaliation events and 2) Target-Selection events.

Response/Retaliation events are defined as offenses that are first triggered by something other than bias, but it is bias that exacerbates the incident and fuels the crime that ultimately occurs. For example, if Motorist A and Motorist B become engaged in a dispute over a parking space which develops into an assault, one could argue that the fight was the result of an argument and not bias. However, if the non-criminal argument escalates into a fight because of the real or perceived differences between the drivers, race, ethnicity, or sexual orientation, for example, this event can now be classified as a hate crime. The question for police officers would be this: Would the incident (the fight not the initial argument) have occurred if the two motorists were from the same group, such as white males? If the answer is "no," that it would not have occurred, then one can say that the incident is motivated "in part" by bias.

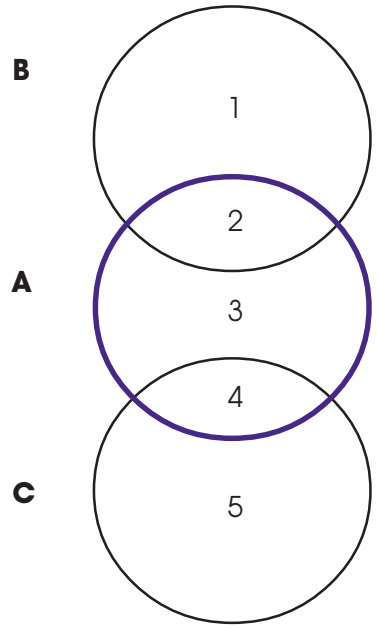
The second type of ambiguous hate crime comes from Target Selection events, which involve perpetrators who are motivated to commit some act – criminal or non-criminal – which itself does not involve bias. For example, a burglar may select a middle-class neighborhood to target during the daytime because he or she thinks residents are more likely to be at work and not at home. The decision to target this neighborhood is derived from a rational thought process and careful consideration. The offender's selection of this

neighborhood is not influenced by any bias against race, religion, sexual orientation, etc. However, there are other criminal offenses where the selection of a target does include bias against the victim group. For example, a criminal offender may be motivated to commit robbery because of a drug habit or because of mounting bills. If this person selects as his or her target members of certain groups, such as gay men coming out of a bar or members of a Jewish group leaving a synagogue because of bias against the group, this act could also be classified as a hate crime. The logic of the discussion above is depicted in Figure 4.

The overlapping circles in Figure 4 represent sets of crime types that are recorded in the police database. Set A includes all police reports that fit the FBI definition of hate crime or the definition provided by the West Virginia criminal code. Set B includes all police reports that involve Response/Retaliation events and Set C includes all reports of Target Selection events. Notice that some, but not all, of the Response/Retaliation events (Set B) involve hate crimes (Set A). Also notice that some of the Target Selection events (Set C) are also hate crimes (Set A). Our work in uncovering hate crimes in the police department records division involved locating reports that would fit into regions 2, 3, and 4 of Figure 4. Police reports that fit into Region 2 are hate crimes that were motivated by some other event but escalated into a crime because of bias. Reports that fit into Region 3 are those crimes that are motivated in whole by bias. Region 4 contains police reports that described the targeting of a victim because of bias but not necessarily as the primary motivation for the crime.

Examples of criminal offenses that fit into the five regions of Figure 4 are described in Figure 5.

Figure 4. Typology of Hate Crime Reports



A = The set of crime reports that fit the FBI definition of hate crime or hate crime defined by state law.

B = The set of crime reports that describe Response/ Retaliation Events.

C = The set of crime reports that describe Target Selection Events.

Other Factors Influencing Reports of Hate Crime

In addition to the problems related to defining hate crime, prior studies have shed light on social forces that influence hate crime reporting by police officers and law enforcement agencies. To an individual officer, these forces can come externally (e.g. organizational climate) and internally (e.g. personal prejudices). In the same manner, an entire agency's handling of hate crime reporting is also shaped by forces from both inside and outside the organization.

Organizational Factors. The culture and norms of a police department are perhaps the strongest influences on individual officer behavior, as they mold the actions of everyone within the organization, from patrol officers to detectives to upper-level management. Organizational norms set a standard for how members of a police force are supposed to act on the job (Bell, 2002). Furthermore, organizational norms guide police on how to treat victims, offenders and witnesses; prioritize cases; and decide which charges to seek in a case (Ericson, 1981). As outlined in prior research, organizational elements can sway hate crime reporting in either direction. Some police departments provide an environment conducive for accurate hate crime reporting: they commit resources to the cause, establish special bias crime units, and urge officers to treat possible hate crime cases with care and precision. Other agencies, on the other hand, hold a more negative view of hate crime reporting. These organizations believe that interpreting and applying hate crime laws creates an additional burden for police on the scene, or they erroneously label hate crime

Figure 5. Types of Police Reports Relating to Bias Crime Reporting

Type of Offense	Example
Response/Retaliation crime that is not bias-motivated.	A white male punches another white male after their cars collide in a traffic accident. No inter-group bias is involved.
Response/Retaliation crime that contains levels of ambiguous bias but could fit either the FBI definition or WV hate crime statute.	A white male punches a black male after their cars collide in a traffic accident. The white male shouts racial slurs at the black male.
Crimes motivated wholly by bias and that undoubtedly meet the FBI definition or WV hate crime statute.	A cross is burned on the front lawn of a black family who just moved into a white neighborhood.
Target Selection crime that contains levels of ambiguous bias but could fit either the FBI definition or WV hate crime statute.	A group of males robs patrons leaving a gay bar. The offenders target this group because they think they are less likely to report the crime to police because they were at a gay bar.
Target Selection crime that is not bias motivated.	A group of males burglarize a middle-class neighborhood during the day because they believe residents are likely to be working and not at home.

policies as unjust, believing that they give special protection only to minorities (McDevitt, Cronin, Balboni, Farrell, Nolan, and Weiss, 2003).

In articles published in 1999 and 2002, Nolan and Akiyama assessed the organizational climate for hate crime reporting. Through focus group interviews in police departments across the country, the authors identified a list of forces – categorized as "encouragers" and "discouragers" – that affect hate crime reporting on both the agency level and the individual level. As the names suggest, encouragers support participation in hate crime reporting while discouragers dissuade it. Agency-level encouragers included an organization's belief that hate crime reporting would improve police/community relations, enhance their ability to assess intergroup tensions in the community, and be the right thing to do politically and morally. Agencies that were considered good reporters viewed encouragers as top priorities of their organization, more so than agencies that were considered non-reporters of hate crime.

Discouragers, meanwhile, included an agency's belief that hate crime reporting is not a priority of local government, that their organization does not provide sufficient staff or technology to deal with hate crime data, and that hate crime reporting is not "real" police work.

There is also a perception that reporting hate crimes results in negative publicity, supports the agendas of gay and minority groups (which can be seen as a negative thing), and makes things worse for communities. Agencies that were considered non-reporters in the previous study were more likely to see discouragers in play at their respective organizations compared to agencies that were viewed as good reporters. This type of organizational mindset was present in Bell's (2002) study of a police department's bias crime unit in a large city (with a population between 500,000 and 900,000). She found that officers who were assigned to this special bias crime unit were often ridiculed and looked down upon as not being "normal cops."

Though not the sole driving force, proper training plays a part in casting an organization's hate crime reporting efforts in a positive light. The development of the national hate crime data collection system corresponded with FBI training programs for law enforcement agencies across the country. The Bureau developed training modules and sessions that went beyond the basics of bias crime reporting by delving

into the causes of prejudice and discrimination. Police are found less likely to participate in reporting if they are not trained properly (Nolan and Akiyama, 2002). Despite the FBI's efforts, some departments met training programs and new hate crime laws with resistance. An organization's dedication to hate crime reporting depends on its commitment to training its own officers.

It is evident that leadership plays a crucial role in establishing the norms and priorities of a police department. The same perspectives held by police administrators regarding hate crime reporting trickles down throughout the entire agency.

Personal Factors. In reality, like everyone else, police officers are imperfect. They can be subjective stewards of the criminal justice system who may foster personal prejudices. These internal beliefs can spill out into the field and be just as influential as organizational variables. In her study, Bell (2002) raised the issue that patrol officers unavoidably bring their own biases into the world of policing hate crimes. She found that some officers did not believe in enforcing bias crime laws against white offenders. This personal belief gave those officers the authority to effectively nullify hate crime law. Individual officers also expect personal incentives or disincentives that accompany whether or not they participate in hate crime reporting (Nolan and Akiyama, 2002; Shively, McDevitt, Cronin, and Balboni, 2001).

In addition to gauging departments' responses, Nolan and Akiyama surveyed officers on "encouragers" and "discouragers" that affected them on an individual level. The authors concluded that officers from the good reporting agencies saw personal incentives if they treated hate crime investigations seriously while officers from non-reporting agencies saw disincentives. Discouraging internal forces – those that affected an individual officer's participation in reporting— included his or her personal feelings that hate crimes should not be treated as "special," that hate crime laws run counter to their personal beliefs, and that enforcing hate crime laws seems more like "social work" than police work. Respondents in the study also expressed little interest in promoting the political agenda of minority groups, such as the gay community. Bell further noted in her study that minority groups historically distrust police and that some officers are reluctant to enforce anti-gay crimes. Much literature has tackled the severed relationships between police and minority groups, which adds another layer to possible police prejudice influencing hate crime reporting.

Classification Error and the "Statistical Accuracy" of Hate Crime Reports

It is understood among those who routinely utilize UCR data that it is a "statistical program," not an actual accounting of all crimes. Even so, it remains a valuable resource for both researchers and the police in their efforts to understand the nature and extent of crime. The value of the UCR is then not contingent on the FBI or state UCR program officials eliminating all errors but rather understanding and measuring error.

It is clear that not all crimes are reported to the police and therefore are not accounted for in official reporting. This type of error occurs on the front end of the process and is the result of crimes simply not being reported either by victims or the police. Prior research has focused on the reasons behind these types of errors. However, the concern addressed by this study is classification error, which is less understood.

Classification error refers to the situation where police officers do record the facts of the incident, but misclassify the crime type. More specifically, the focus here is misclassification due to the failure to distinguish a hate crime from any other crime. Behaviorally, hate crimes are "regular" crimes such as burglary or assault. What distinguishes these crimes as "hate crimes" is the bias that motivates the crime or the selection of the victim (FBI, 1997). Therefore, a hate crime can be classified incorrectly if the bias motivation is missed or even ignored by the police when reporting the incident. Classification errors such as these can then have a substantial impact on the statistical accuracy and interpretation of UCR crime estimates.

Statistical accuracy refers to the errors found in the crime totals after all crime types have been examined and offsetting misclassifications have been considered. Some misclassifications will result in overcounts of UCR crimes, while others result in undercounts. The correct UCR number can be obtained by considering the canceling effect of the two types of errors, overcounts and undercounts.

Methodology

West Virginia Incident-Based Reporting System (WVIBRS) data files for 2008 were used to partition records, calculate sample size, and randomly select records for review. These data files for 2008 (as well as prior years) were first used to examine the number of hate crimes reported annually and which cities/counties were reporting them. In addition, the specific types of crimes reported as having hate or bias factors were reviewed. In WVIBRS, hate crimes are identified by a separate variable used to indicate that an offender's actions were motivated, in whole or in part, by bias against a specific group. Incidents involving any offense type can include bias motivation.

Records in WVIBRS were partitioned into four distinct categories. First, all records that included the hate crime indicator were separated out (i.e, Hate Crime). The research team planned to review the total population of these records. Second, assaults, robberies, burglary/breaking and entering, and destruction of property/vandalism incidents appeared to be the most likely Group A offenses to have hate or bias identified as a motivating factor. Therefore, these offenses were grouped separately from all other Group A types of crimes reported in WVIBRS (i.e, Group A). The remaining Group A offenses formed the third category (i.e., Other A). Lastly, Group B offenses (which are only recorded in WVIBRS when an arrest is made and thus do not include the hate crime indicator) comprised the final group of offenses (i.e., Group B). In addition, the research team further felt it was important to obtain and review a sample of cases that law enforcement considered as general information incidents and unfounded reports. It was necessary to examine these incidents and reports to determine whether some hate crimes were being missed in reporting. Since these records are not included in WVIBRS data files, however, it was not possible to presample cases in this category.

Of the 92,939 incidents reported to law enforcement in 2008 only 60 were identified as involving a hate crime. Due to the relatively low volume but widespread geographical distribution of reported hate crimes in West Virginia, researchers decided to limit the study to four law enforcement agencies in the state. The agencies selected include two larger municipal police departments and the sheriff's departments for the counties in which the cities are located. These agencies represent two of the more heterogeneous communities in West Virginia. From the partitioned data files, a random selection of cases was chosen for review. Researchers provided the agencies with a list of incident numbers and scheduled on-site visits to review the paper case files. Two of the agencies elected to provide copies of the reports which the researchers reviewed as a team at their office.

Calculating Sample Size

The sample size was selected based on Equation 1 shown below.

$$\textbf{(1)} \qquad \frac{k^2NPQ}{k^2PQ + NE^2}$$

where,

k = confidence level (1.96 represents 95.0% confidence)

P = estimated proportion of hate crimes in the stratum population

Q = (1 - P)

E = desired precision (0.03).

In order to select the sample from each stratum, the P value was estimated based on prior experience with the reporting of hate crimes. P values were set as follows,

Group A, P = 0.2

Other Group A, P = 0.1

Group B, P = 0.05

Figure 6 illustrates the sample size calculated for each offense category. The sample breakdown is shown for each selected agency along with the total reported in each offense category.

Selection and Review of Sampled Records

The 2008 WVIBRS incident-level data file was imported into SPSS and used to generate the random sample of cases for review. A variable was created to separate the Group A incidents into two categories based on the most serious reported incident offense. Incidents involving a reported hate crime were flagged as such so that all of these cases could be reviewed. A separate file of Group B arrest reports was also utilized to randomly select cases from this category for review.

Figure 6. Population and Sample Sizes by Offense Category

	Municipal Agency A		Municipal Agency B		County Agency A		County Agency B		Total
Stratum	**N**	**n**	**N**	**n**	**N**	**n**	**N**	**n**	**Sample**
Group A	3,003	207	3,095	203	2,339	160	875	66	**636**
Other A	3,479	149	2,832	121	1,692	67	904	31	**368**
Group B	5,073	112	1,347	21	1,966	42	1,359	25	**200**
Hate Crime	4	4	8	8	6	6	0	0	**18**
General Information	108	43	72	25	0	0	177	34	**102**
Total	11,667	515	7,354	378	6,003	275	3,315	156	**1,324**

Note: The population of general information reports was approximated by and the sample was selected by agency personnel. One county agency indicated that they did not have this type of report. N = population, n = sampled records.

Based on the determined sample size required, a random sample of cases reported by the chosen agencies was selected. A list of the selected cases identified by incident or arrest number was generated. The lists were then forwarded to each agency's records division supervisor where, hardcopies of the identified cases were manually pulled for the research team to review. A team of eight reviewers spent three days reading and assessing files on site and another day reviewing records at their offices.

In reviewing each record, the definition of hate or bias crime as established by the FBI as well as the state code was applied. Prior to the record review, team members participated in a one-day training focusing on the definitions and bias indicators. James Nolan provided the team with information on understanding and clarifying ambiguities in bias crime classification.

To ensure a high level of reliability between reviewers, a systematic procedure for the assessment of each record was established. From the cases selected for review, 10% from each category were further randomly assigned to receive a second review from a different team member. In addition, for those records where a classification error was suspected, the team members worked together to come to a consensus on the recommended classification. Since only 18 cases in the sample were indicated by the police agencies as hate crimes, *all* of these cases were reviewed by the full research team.

During the on-site review, the agencies were also asked to randomly pull a sample of cases from their general incident/information files to be assessed. These cases represent situations where law enforcement responded but determined that no criminal offense had been committed. A total of 102 general information records were randomly selected by the agencies for review. In addition, the researchers inquired about unfounded case files (i.e., reported crimes that were subsequently found to be false). While the agencies indicated that no such cases were available for review, the researchers discovered that some of these cases were actually present in the sampled Group A and Other A cases. The 33 cases in the sample that reviewers considered unfounded were deemed to be overcounts and will be discussed later in the results.

Key Terms and Definitions

The definitions of the crime categories as established for the purposes of this study are included below along with statistical terms used frequently throughout this report.

Group A Offenses. Offenses most likely to occur in situations where bias is a motivating factor. These offenses include all assaults (aggravated, simple, and intimidation), burglary/breaking and entering, destruction/damage/vandalism of property, and robbery.

Other Group A Offenses. All other crimes against person, property, or society for which incident reports are required to be submitted to NIBRS/WVIBRS. These offenses include arson, bribery, counterfeiting/forgery, drug/narcotic offenses, embezzlement, extortion/blackmail, fraud offenses, gambling offenses, homicide offenses, kidnapping/abduction, larceny/theft offenses, motor vehicle theft, pornography/obscene material, prostitution, forcible and nonforcible sex offenses, stolen property offenses, and weapon law violations.

Group B Offenses. Crimes that are only reported to NIBRS/WVIBRS when an official arrest is made. The offenses included are bad checks, curfew/loitering/vagrancy violations, disorderly conduct, driving under the influence, drunkenness, nonviolent family offenses, liquor law violations, peeping tom, runaway, trespassing, and all other offenses not defined as Group A.

General Incidents. All reports filed by the police for noncriminal matters, such as suspicious person investigations, false burglary alarms, and community problems/disputes.

Hate Crimes. As defined by the FBI, "a criminal offense committed against a person or property which is motivated, in whole or in part, by the offender's bias against a race, religion, disability, ethnic/national origin group, or sexual orientation group." West Virginia also recognizes gender and political affiliation as protected classes against bias, and these were considered for the present study.

Hate Incident. While not a criminal offense, an incident that is reported to law enforcement and involves characteristics of bias against a race, religion, disability, ethnic/national group, or sexual orientation group. West Virginia also recognizes gender and political affiliation as protected classes against bias, and these were considered for the present study.

Unfounded. Crimes that were reported to law enforcement but were subsequently determined by police to be false or baseless.

Confidence Intervals. The interval of values surrounding the point estimate in which researchers can be confident that the true population parameter (e.g., the number of crimes) falls.

Point Estimate. A statistic provided without indicating a range of error. The best guess of the true number of crimes in each crime category in the population under study.

Overcounts. When reports in crime category X are examined, overcounts represent reports that should have actually been in some other category Y. These reports are deemed overcounts of category X.

Undercounts. When reports that should have been in category X are found in another category Y. The reports result in an undercount of category X.

Statistical Definition. The UCR definition of a crime.

Criminal Definition. The criminal definition of a crime as written in state code.

Results

The results of this study focus on the statistical accuracy of bias-related incidents reported by the selected law enforcement agencies in West Virginia. As described previously, the findings center on the degree to which offsetting overcounts and undercounts were found in the classification of hate crimes. Emphasis is placed on the presence of bias crime indicators in police reports across six distinct categories: Group A offenses (A), Group A Hate Crimes (AHC), Other Group A offenses (OA), Other Group A Hate Crimes (OAHC), Group B arrests (B), and General Incidents (GI). An assessment of classification error found

Figure 7. Matrix of Overcounts and Undercounts

Original Police Classification	Recommended Reviewer Classification: A	AHC	OA	OAHC	B	GI	GIHI	U	Total	Overcounts
A	**605**	3	2	0	0	2	0	31	643	38
	8,762	43	29	0	0	29	0	449	9,312	550
AHC	11	**5**	0	0	0	0	0	0	16	11
	11	**5**	0	0	0	0	0	0	16	11
OA	1	0	**357**	0	0	0	0	2	360	3
	25	0	**8,833**	0	0	0	0	49	8,907	74
OAHC	0	0	2	**0**	0	0	0	0	2	2
	0	0	2	**0**	0	0	0	0	2	2
B	0	0	0	0	**185**	0	0	0	185	0
	0	0	0	0	**9,745**	0	0	0	9,745	0
GI	3	1	8	0	3	**86**	1	0	102	16
	3	1	8	0	3	**86**	1	0	102	16
GIHI	0	0	0	0	0	0	**0**	0	0	0
	0	0	0	0	0	0	**0**	0	0	0
U	0	0	0	0	0	0	0	**0**	0	0
	0	0	0	0	0	0	0	**0**	0	0
Total	620	9	369	0	188	88	1	33	**1,308**	
	8,801	49	8,872	0	9,748	115	1	498	**28,084**	
Undercounts	15	4	12	0	3	2	1	33		70
	39	44	39	0	3	29	1	498		653

within each category is also provided. In addition, focus groups with law enforcement officers provide insight into the thought process involved with classifying hate crimes.

Assessing Overcounts and Undercounts

Figure 7 illustrates the overcounts and undercounts found in each crime category in a matrix format. Each row depicts the *original* classification assigned to the incident by the law enforcement agency. The columns reflect the classification *recommended* based on the reviewers assessment using bias crime definitions and indicators.

Each cell of the matrix contains two numbers representing the sample (top) and the population estimate (bottom). For example, the first cell representing the intersection of original "A" and recommended "A" (denoted as A|A) contains the numbers 605 and 8,762. This indicates that 605 of the Group A records sampled were confirmed as Group A incidents. The number 8,762 is then the estimate of the number of Group A incidents in the population of reports that were reported accurately by the police. The remaining cells along the main diagonal of the matrix (shown in bold) therefore indicate the number of records in the sample where reviewers agreed with the police (top) and the estimate of accurately classified records (bottom) for each crime category.

Finally, the last *column* of Figure 7 provides the overcounts in the sample and the population estimate of overcounts by category, while the bottom *row* denotes the undercounts in the sample and the population estimate of undercounts by category. The intersection of these two illustrates that there were 70 records out of the 1,308 in the sample that were identified as classification errors when considering both overcounts and undercounts. Based on this study sample, it was then estimated that 653 classification errors were contained in the population of 28,084 records.

In analyzing the errors in bias-motivated incidents individually, 5 undercounting errors and 13 overcounting errors were identified in the study sample. That is, 5 records which were *not* originally identified as hate crimes/incidents by the police were determined by the reviewers to have sufficient indicators to be counted as hate crimes/incidents. On the other hand, 13 records which the police *did* indicate as bias motivated were not judged to be hate crimes/incidents by the review team (see Figure 7).

The error rates in the sample are then applied to the population of reports in each category to establish point estimates of the actual number of crimes. For example, the 4 undercounting errors in the "AHC" category predicts 44 undercounted hate crimes in the population of "AHC" reports (see bottom cell of column "AHC"). Most of the undercounted hate crimes come from the "A" crime category, cell A|AHC. In addition, the 1 shown in cell GI|AHC indicates that reviewers believed that one of the reports originally classified as general information was actually a Group A hate crime. The final undercount is shown in the cell GI|GIHI which indicates that while it did not rise to the level of a hate crime it was a racially charged non-criminal incident (denoted as GIHI). The overcounts are shown in cells AHC|A (11) and OAHC|OA (2).

The remaining cells above and below the main diagonal contain zeros when no errors were found across the intersecting categories. Since the focus of this study involves examining errors in the identification and classification of hate crimes, reviewers were less concerned with verifying the specific offense recorded. While these types of errors were less likely to show up in our analysis of broad offense categories, some are recorded in the matrix but will not be discussed in detail.

It is important, however, to note that *33 cases coded as Group A or Other A by police were judged to be unfounded by the research team*. This is another source of error that the research team did not set out to discover, but uncovered through the process of reviewing records. In the end, this type of error resulted in a large number of estimated overcounts in the population and is likely to inflate estimates of crime substantially – if it is widespread across police agencies.

While these 33 cases do *not* contribute to the error in *hate crime* reporting, (because the errors are not due to bias motivation), they do represent a fairly large number of overcounts in the sample. In this particular case such error resulted in an estimate of 498 records in the population of Group A and Other A crimes that may actually be unfounded (denoted as U). Many of these cases were originally recorded as *unfounded* on the paper copies of the police records; however, they were mistakenly counted as crimes in the WVIBRS electronic files from which the sample was selected. Based on our review, these cases often involved situations where police arrived at a domestic incident but no signs of physical violence were present. Both parties

Figure 8. Case Narratives of Hate Crime Undercounts

Case 1:
A black woman walking home from a sub shop encountered a white man walking two large dogs. One dog had a leash; the other did not. The dog with no leash got near the black woman, who asked its owner to keep the dog away from her. She asked again after the dog owner never replied. This time the white man responded, "what are you going to do, nigger?" The black woman replied, "I'll kick your dog." The white man approached her and "body bumped" her before running away.

Case 2:
A black male juvenile sees a 47 year old white woman walking her dog and invites her onto his porch. After stepping onto the porch, the white woman was grabbed and pushed by the black juvenile. The boy then struck her on the face "while yelling racial comments" at her. The woman ran from the porch but was chased by the juvenile, who threw rocks at her. When interviewed by police, the woman said she did not know the male or his exact address. The responding officer searched for the suspect but could not find him. No complaint was signed due to lack of suspect information.

Case 3:
A mother tried breaking into her daughter's home by tearing through a locked screen door. The mother came to retrieve a Jeep given to her daughter as a graduation present. An argument then ensued between the two, and at some point, the mother grabbed her daughter by the throat. According to the police report, the mother was angry because her white daughter was pregnant by her black boyfriend.

Case 4:
Someone left a note in a hotel room that said, "any female found in this room will be raped." No other information was available; therefore, police filed a general incident report.

indicated that it was only an argument and did not wish to pursue charges or seek protection. In this situation, one party often voluntarily leaves the scene to end the confrontation. In the end, this type of scenario often resulted in the incident being recorded as unfounded, unless additional information came forth.

Illustrative Examples of Undercounts. To further examine the specifics of the cases in which reviewers identified hate crime undercounts, case narratives depicting the *four* incidents are presented in Figure 8. Recall, undercounts represent those cases where law enforcement did *not* recognize and record an incident as a possible hate crime. While perhaps not as clear cut as the high profile cases mentioned earlier, the research team believes that each of these cases contains sufficient details to indicate the possibility of bias. As a result, these cases appear to be at least "in part" motivated by bias and thus meet the statistical definition for reporting as hate crimes.

In Case 1, the responding officer's narrative noted the difference in races between the two persons involved. In addition, along with the physical assault, the dog owner used a racial slur for the black victim. Similarly in Case 2, the difference in race is noted and racial slurs are again present. Moreover, no additional information is provided to establish any alternative motive. Although the victim and offender are both white in Case 3, it suggests that race was a motivating factor for the assault. The report states that the mother's anger with her daughter is the result of her being pregnant by a black boyfriend. Each of these three cases were originally coded as Group A incidents with no indicator of bias and were thus considered undercounts of bias crimes by the research team.

Based on a review of Case 4, which was found in general incident reports, the researchers determined that leaving a threatening note could constitute intimidation. Also, this threat is not specific to one person but instead targets all females. The research team reclassified this case as a Group A incident rather than simply a general incident because a crime was involved, even though no suspect was reported. Furthermore, this case does fit the definition of a bias-motivated act since gender is a protected class under WV Code.

In addition to the four case narratives involving crimes in Figure 8, one more undercount was identified as a hate incident (refer to GI|GIHI, Figure 7). Because this incident

Figure 9. Case Narratives of Accurately Reported Hate Crimes

Case 1:
A white male is found laying on the front porch of a house with his pants around his knees, blood coming from his nose and the top of his head, and a small laceration on his arm. The male indicated he was sitting on his front porch when he heard something to the rear of the house. He went to check it out and saw three males of unknown race wearing masks. The three males saw him, called him by name, and came running at him yelling faggot and queer. One of them had a baseball bat and struck the victim several times. The other male punched him in the face and head while the third assailant had a knife or broken bottle that was used to cut the victim's arm. The victim was eventually knocked out and when he woke up he was being sexually assaulted by one of the males. The victim stated that the males kept calling him faggot and queer during the assault. Afterwards the three males fled from the backyard. An aluminum baseball bat was found under the victim's van.

Case 2:
A white female juvenile stated that she was sitting with her friend on the sidewalk waiting for her bus when a group of black female juveniles walked up behind her and acted like they were going to punch her. The victim's friend told her the girls were going to jump her, he noticed them plotting it out. The victim then asked one of the girls why she was getting ready to hit her. Several of the black girls then jumped the victim punching her several times in the face and dragging her across the lot. The girls then fled toward a nearby cinema.

Case 3:
While walking home a black male noticed some people across the creek with a bonfire and yelled "hey let's party." The subjects replied with a disparaging remark and racial slur. The black male went into his house and told his father what happened. When the father came out of the house a white male came down the creek yelling "I'll burn your house down nigger." This male was also throwing rocks at the house. Later, the black male reported a broken window in his truck which he believes happened during the altercation.

Case 4:
While attempting to leave her residence with a friend, a black female was blocked in so that she could not leave by two white males in a pickup truck. The white males harassed the female and her friend using racial slurs for several minutes before allowing them to leave. The black female then ran to her sister's house and hid behind the fence. She stated that she feared the men would hurt her and her family because they are black.

Case 5:
An 18 year-old female stated that she was sitting on a bench outside a supermarket with her girlfriend when her mother, who she has little contact with, exited the store. When the mother noticed another female, who the girls reside with, she became irate. The young female stepped between her mother and the other female and asked her mother to back off. The mother pushed her daughter away, called her a "stupid lesbian" and the two engaged in a physical fight. The mother then fled the scene yelling threats at her daughter.

does not impact the statistical accuracy of crime reports, it is only briefly mentioned here. It involved a neighborhood disturbance with a man who frequently used racial slurs and, on this occasion, made inappropriate sexual comments about a group of teenage black females. While this man's comments and actions were racially charged, they were not criminal. Therefore, this case was classified only as a hate incident.

Illustrative Examples of Accurately Reported Hate Crimes. Five other hate crime reports were reviewed as part of this study and reviewers agreed with the police that they were bias-related crimes. A brief description of these cases is provided in Figure 9. While Case 1 is more detailed and the bias indicators may be more pronounced, Case 2 is not that different from the previous four undercount cases discussed in Figure 8. The last three case examples contain language suggesting bias against particular groups.

As illustrated in Figure 9, a white male victim is violently assaulted both physically and sexually in Case 1. The assailants were wearing masks but apparently knew the victim since they called him by name. During the assault they continued using offensive names which implied bias toward the victim's sexual orientation. From the available details it appears that the victim was targeted for this reason. Hence, researchers agreed with the law enforcement agency that this was a hate crime. It is interesting to note that sexual orientation is *not* one of the protected classes under WV Code; however, the police correctly recognized this case as a hate crime.

Similar to three of the four crimes previously described as undercounts, Case 2 identifies the different races of the victim and her offenders. A white juvenile girl is jumped and assaulted by a group of black juvenile girls for unknown reasons. The victim did not indicate or appear to know the girls. Given the available information and lack of other motivation, the incident seems to be at least partially motivated by race and is considered a hate crime both by the police and the research team.

Cases 3 and 4 were reported by the police as anti-black hate crimes. In both cases, the black victims appear to be harassed by white assailants who repeatedly use the "N" word to address their victims. The black male's property was damaged in Case 3, while the black female in Case 4 was assaulted. The research team agreed with the police that both cases are hate crimes.

In Case 5, the domestic assault appears to be the result of a mother's intolerance for her daughter's sexual preference and living situation. Even though the mother is not the legal guardian and has had little contact with her daughter, she initiates an altercation upon seeing the girls together. The mother's language confirms her disapproval of her daughter's lifestyle and indicates that her actions are bias motivated.

Explanation of Overcounts. In terms of overcounts, a total of 13 records originally classified by police as hate crimes were *reclassified* by the research team. In each of these incidents, no evidence of bias motivation was found in the officer's narrative. It is important to note, however, that further investigation into these cases beyond the examination of incident reports and narratives may have led to other pertinent information not available at the time. Likewise, it is possible that these errors might have simply been the result of data entry mistakes. In fact, upon review of the original written reports, the majority did *not* have the box checked for the "hate crime indicator" but in the WVIBRS data file they were incorrectly recorded as such. Regardless, they were considered overcounts and were reclassified to Group A or Other Group A incidents. These overcounts are identified in Figure 7, and classified accordingly. Illustrative examples of these overcounts would provide no meaningful information since they simply describe crimes—with the absence of any description of bias or hate motivation.

Impact of Classification Error on Statistical Accuracy of Crime Reports

Using the data assembled in the "Matrix of Overcounts and Undercounts" shown in Figure 7, we are able to combine all crime categories and assess the overall impact of classification error on statistical accuracy. Recall that "statistical accuracy" refers to the errors found in the crime totals after all crime types have been examined *and* offsetting misclassifications (i.e., undercounts and overcounts) have been taken into account. The findings are presented in Figure 10.

Statistical accuracy is calculated in two steps. The first step involves estimating the true number of crimes in a particular crime category. This actual crime total is denoted by T_i. In order to estimate the number of hate crimes in this study, we first calculated estimates for all eight categories in the database (i.e., A, AHC, OA, OAHC, B, GI, GIHI, and

Figure 10. Crime Estimates, Error Rates, and Confidence Intervals

Crime Category	Reported	Estimate	Error	Low	High
Group A (A)	9,312	8,801	5.81%	8,506	9,096
Group A Hate Crime (AHC)	16	49	-67.35%	1	97
Other Group A (OA)	8,907	8,872	0.39%	8,718	9,026
Other Group A Hate Crime (OAHC)	2	0	N/A	N/A	N/A
Group B (B)	9,745	9,748	-0.03%	N/A	N/A
General Information (GI)	102	115	-11.30%	76	154
General Information Hate Incident (GIHI)	0	1	N/A	N/A	N/A
Unfounded (U)	0	498	N/A	281	715
Total	**28,084**	**28,084**	0.000%		

U). For example, in order to estimate the statistical accuracy of the Group A Hate Crimes, (AHC), we first established the true number of AHC crimes in the population of reports, denoted here as T_{AHC}. T_{AHC} is calculated according to Equation 2 below.

(2) $T_{AHC} = N_{AHC}$ - (overcounts) + (undercounts) =

$T_{AHC} = 16 - (11) + (3/643)9{,}312 + 1 =$

$T_{AHC} = 16 - (11) + (43 + 1) = 49$

N_{AHC} is the number of AHC originally reported in the population.

By Equation 2, the true number of crimes that fit the Group A Hate Crime (AHC) category is 49. Originally only 16 were reported, indicating that Group A Hate Crimes were undercounted by -67.3%

Confidence intervals for each estimate were calculated according to Equation 3 below.

(3) $1.96\sqrt{(N-n/N)\ (PQ/n-1)}$, where

P_i = percent error found in the sample
$Q_i = 1 - P$
n_i = sample size
N_i = total reports in the population of crime category i

The Impact of Classification Error on Hate Crime Totals

Figure 10 provides a summary of the point estimates and confidence intervals for each crime category. The "reported" column illustrates the number of reported incidents in each category. The "estimate" column provides the point estimate of crimes based on the review of sampled records. This statistical error percentage is reported in the "error" column. Negative percent error indicates an undercount of crimes in a given category. Meanwhile, positive percent error is indicative of overcounts in a given crime type. In addition, the "high" and "low" columns provide the upper and lower bounds of the confidence interval.

Despite the small number of errors found in the 1,308 reports sampled, the resulting error estimates can become quite pronounced. As shown in Figure 10, a vast majority of the error found in this study involved the undercounting of Group A Hate Crimes (-67.35%). A total of 3 crimes originally classified as Group A (A) and 1 originally classified as a General Incident (GI), were reviewed by the researchers and found to actually be hate crimes. As a result, the findings estimate that 49 incidents from this category could be classified as Group A Hate Crimes (AHC) in the population. The findings also suggest an overall undercounting error in the General Incident category by -11.30%. While 102 General Incident reports were

reviewed, it is estimated that 115 could actually exist in the population. These results suggest that most of the error in crime reporting is predominately from undercounting given the negative signs. This finding is consistent with research that suggests hate crimes often go undetected by law enforcement.

The 13 overcounts originally classified as hate crimes then contribute to the error estimates in the Group A and Other Group A categories shown in Figure 10. Overall the effect of the overcounts is not as great due to the small number of records making up the hate crime population. Additionally, because the true number of reported Unfounded crimes is not know (indicated as none by these agencies), the error in this category can not be calculated. However, the 33 cases found in Group A and Other Group A crimes, that were reclassified as Unfounded, predict that 498 may exist in the population. This implies that Unfounded cases are *undercounted* to some extent. More importantly, these Unfounded cases appear to be contributing to the *overcounting* of actual crimes.

Officer Explanations for Classification Error

After all records were reviewed and classified by the research team, a focus group was held with officers from one of the law enforcement agencies involved in the study. The purpose of the focus group was to obtain insight into the thought processes officers adhere to specifically regarding hate crime policing. The interview lasted over one hour and included four patrol officers and a lieutenant with varying levels of experience. In addition to some general questions, the officers were presented with some of the case scenarios reviewed by the research team.

Several themes stood out during the discussion with the officers. One overriding theme seemed to be that if they encountered a hate crime, it would be obvious and they would easily recognize it as such. While they acknowledged that hate crimes do occur everywhere, these officers did not think that they had seen one. The consensus seemed to be that "when you do see a hate crime, it is the type that makes the national headlines." This is what the officers referred to as a "clear-cut case."

From the perspective of officers, lack of a relationship between parties and serious injuries or death were important characteristics for determining whether a case involved a hate crime. They felt that "if the parties did not know each other, then the bias might be the only reason for the crime."

From the Perspectives of Officers...

On the overt nature...
"...when you see a hate crime, it is the type that makes the national headlines...referred to as a 'clear-cut case"

On the importance of relationship...
"...if the parties did not know each other, then the bias might be the only reason for the crime"

On other factors such as location...
"...people just do not go onto a stranger's porch for no reason"

On getting to the truth of the matter...
"...[we] often get one story from the victim and another story from the offender and often the truth falls somewhere between"

On victim statements...
"...even though a victim may believe they were targeted due to a bias, that doesn't prove the offender's motivation"

Furthermore, it was necessary for the relationship to be coupled with serious injury or death. Other indicators for officers involved the presence of graffiti or racial slurs spray painted on someone or at the scene. Cited examples included swastikas, cross-burning, or setting a synagogue on fire.

In their experience, however, these officers seemed to agree that most crimes could be attributed to other factors such as location or even prior dealings gone bad between the parties. For example, in the case of the white female walking her dog who goes onto the porch of the black male juvenile (see Figure 8, Case 2), the officers felt certain that there was more to the story. They believed that she must have had some reason to go there, particularly in the neighborhood where this incident occurred. The officers indicated that "people just do not go onto a stranger's porch for no reason." Likewise, disputes over drugs and/or money were mentioned repeatedly by the officers as instigators for many crimes.

Officers further indicated that one of the biggest hurdles they face in classifying crimes is simply "getting to the truth

of the matter." Officers stated that they "often get one story from the victim and another story from the offender and often the truth falls somewhere in between." For instance, they suggested it is not possible to depend on victim statements by themselves. In terms of indicating whether a crime was motivated by bias, the officers felt they simply could not rely on the victim statements. One officer stated that "even though a victim may believe they were targeted due to a bias, that doesn't prove the offender's motivation." In the case involving the mother whose daughter was pregnant with her black boyfriend's baby (Figure 8, Case 3), this was clearly an example. The officers said "the narrative only provided why the daughter thought her mother was angry." Therefore, the officers rejected this, as well as other cases questioned by the reviewers, since they considered the reports to be victim statements.

In addition, the general consensus was that hate crime incidents were not spontaneous events. In order for a person to be motivated by a bias to commit a crime, they believed that "the crime would have to be premeditated" – for example, if a person looked for someone to assault simply because they are black. In the case of the black female who was called a racial slur by a white male walking his dogs, the officers said "the slur was probably out of anger over the confrontation" (Figure 8, Case 1). They didn't believe that the man simply went walking with the intent to call a black person a disparaging name.

It should be noted that participants in the focus group were mainly frontline officers taking initial reports and ensuring safety at the scene. The officers routinely stated that such incidents are followed-up on by the Criminal Investigative Division where additional inquiry would take place. Many of the officers stated that "it would be the detectives who would obtain the additional information, if it existed, to classify the incident as a hate crime." Moreover, officers seemed to suggest that the decision to classify an incident as bias-motivated seemed somewhat outcome based. That is, several agreed that "if the victim is not willing to cooperate or to ensure that charges will be pressed, it isn't important to record it." In their experience, as one officer put it, "suspects are often back on the street before we can even get the paperwork completed."

When asked whether additional training would help officers identify hate crimes, the group did not believe it was necessary. Generally, this group of officers felt that "if a hate crime occurred, it would be easy to recognize, just common sense." All of the officers indicated that they had received some training related to hate crimes at the academy. The Lieutenant reported that "some hate crime training had been done probably 15 years ago." In their day-to-day dealings with the public, many of the officers stated they "hear racial slurs all the time but that doesn't make it a hate crime." Thus, officers stated they "do not consider most situations involving racial slurs as hate crimes but rather just the way people talk."

On the whole, the officers seemed to understand that hate crimes are a special category and believed that they should be treated as such. They agreed that "it is the

From the Perspectives of Officers (Continued)...

On hate crimes as spontaneous events...
"...[to be motivated by a bias to commit a crime] the crime would have to be premeditated"

On the role of frontline officers...
"it would be the detectives who would obtain the additional information, if it existed, to classify the incident as a hate crime"

On victim participation and case outcomes...
"...if the victim is not willing to cooperate or press charges, it isn't important to record it....suspects are often back on the street before we can even get the paperwork completed"

On training...
"...if a hate crime occurred, it would be easy to recognize, just common sense....[Lieutenant reported] "some hate crime training had been done probably 15 years ago"

On racial slurs...
"...[we] hear racial slurs all the time but that doesn't make it a hate crime"

On the seriousness of the crime...
"...victims of hate crimes can't really protect themselves from being targeted...hate crime victims can't protect themselves against being female, black, or gay"

motivation that makes these crimes different." One officer noted, and others agreed, that "victims of hate crimes can't really protect themselves from being targeted." A person may be able to lessen their likelihood of becoming a victim of certain crimes by avoiding areas where violence is known to take place. However, "hate crime victims can't protect themselves against being female, black, or gay," one officer stated, for example. In fact, the officers indicated that they would love to check the box (referring to the hate crime indicator on the WVIBRS form noting incident was bias-motivated)— if they were certain. There seemed to be the perception that if an incident involved a hate crime, the outcome may be more harsh. Specifically, a serious hate crime case could be prosecuted in federal court rather than ending up in magistrate court.

Discussion and Conclusions

Utilizing a methodology previously developed by the authors, this study sought to examine the extent to which classification error contributes to inaccuracies in officially reported hate crime statistics. Our approach combined multiple sources of data to explore both the magnitude of error in hate crime statistics and the issues which complicate law enforcement efforts to properly classify such incidents. The results clearly illustrate that classification errors can, and in fact do, impact the statistical accuracy of official hate crime statistics. Our findings showed that the true number of hate crimes in the population is likely much greater than official crime reports suggest.

A vast majority of the error found in this study related to the undercounting of hate crimes in official records. That is, the researchers found evidence of bias in incidents which were never classified as hate crimes by officers. Based on a systematic review of 1,308 incident reports across two city and two county law enforcement agencies, we estimated that Group A Hate Crimes were undercounted by approximately sixty-seven percent (67.35%) in WV's incident-based reporting system. As a result, we estimated that approximately 49 Group A Hate Crimes actually occurred in the study's population of four agencies, rather than the 16 which were originally reported by law enforcement. During this same year *only* 60 hate crimes were reported for the entire state of West Virginia. Clearly, these results have large ramifications for the reporting of hate crime statistics in the state.

The large number of undercounts found in the present study appear to stem from the failure of officers to recognize "bias indicators" when present in a given incident. As the focus group results suggest, this can be a very difficult task for officers. In the classification of such crimes, organizational (i.e., culture and norms of police departments) as well as personal factors (i.e., internalized beliefs of officers) merge and provide a context in which decisions need to be made by officers. When an officer encounters an incident or situation, he or she is asked to ascertain the intentions and motivations of the people involved, and make a judgment as to whether the crime was motivated by bias.

The focus group discussion clearly points to thought processes officers engage in when ascertaining whether a particular incident constitutes a hate crime. And these thought processes can influence the officers' assessment of whether a crime involves hate. In some regards, our focus group officers seemed to indicate that such crimes are "obvious" when they occur. This appears to be tied to the thought that most hate crimes are severe in nature. But as we see from the case narratives, indicators of bias can be slight and the crimes do not have to be serious in order to be motivated by bias, thereby making the classification of hate crimes a very difficult task.

As officers enter these difficult situations, against a backdrop of organizational and personal factors that predispose their thoughts about hate crimes, the inherent difficulties in defining what is (and what is not) a hate crime become more apparent. For instance, officers point to various factors such as the prior relationship between parties, conflicting stories from the victim and offender, as well as whether the victim is likely to participate in the prosecution. Moreover, prior research tells us that some hate crime incidents are more easily defined than others, which further complicates matters. For instance, we found a majority of the undercounted hate crimes to be examples of what previous research has described as Response/Retaliation offenses (i.e., offenses that are first triggered by something other than bias). As discussed previously, these types of hate crimes are more ambiguous and harder to see than crimes motivated in whole by offender bias.

The authors hope that by highlighting the difficulties associated with defining hate crimes, this study will yield useful information for the training of officers on the reporting of hate crimes. It is hoped that this study will bring greater

awareness to the issues associated with the proper interpretation and classification of such crimes. It is clear from our results that such difficulties can lead to errors in classification by law enforcement personnel and lead to less accurate reporting of hate crimes in official statistics.

In examining the issue of statistical accuracy in incident-based data, this study further demonstrated the usefulness of this previously established methodology for exploring errors in specific types of crimes (i.e., hate crime). In utilizing a methodology that "goes to the source" (that is, original agency documents recorded by officers), we were able to directly assess the presence or absence of key bias indicators. However, we recognize that this approach is not without its limitations. In some instances, the "written word" may not truly reflect the "realities" and "dynamics" of the situation, nor the cognitive intentions or motivations of the persons involved. Future research should build on this study's approach by identifying ways to incorporate the perspectives of officers and others involved in the specific incidents. Given that this study illustrated that officers are likely to encounter situations with their biases, future studies should seek ways of involving victims, and if possible, perpetrators of hate crimes. Regardless, our approach offers a readily available and feasible mechanism for allowing law enforcement agencies themselves to monitor reporting.

Finally, it is hoped that this research will contribute to more accurate hate crime reporting in the future. This study provided a sense for how both overcounts and undercounts can occur. On the one hand, it appears that *overcounts* are largely due to data entry mistakes, whereas undercounts appear to be marred in the problems associated with identifying bias when it is present. It is only through the training of law enforcement personnel, and helping officers better recognize the indicators of bias, that we can obtain more accurate statistics on the number of crimes motivated by hate in the United States. Perhaps this will get us a step closer to estimating the true magnitude of these crimes and ultimately lead to new methods for statistically adjusting crime statistics. In this sense, we will gain a more accurate measure for the actual number of hate crimes in the population.

Summary of Key Conclusions....

This study examined the impact of classification error on the statistical accuracy of reported hate crimes.

Two nationally publicized incidents, *the cases of Arthur "J.R." Warren and Megan Williams,* underscore the need to better understand hate crime reporting in West Virginia.

Distinguishing bias-motivated crimes from unbiased crimes remains a difficult practice for law enforcement officials and even experts in the field.

Prior research has identified that both Response/ Retaliation and Target-Selection events pose problems in the identification of hate crimes for officers.

Both organizational (i.e., culture and norms of police departments) and personal (i.e., internalized beliefs of officers) factors contribute to officers' definition of the situation and whether a crime contained hate biases.

Based on this study sample, it was estimated that 653 classification errors were contained in the population of 28,084 records

This study uncovered a large source of error where cases coded as Group A or Other A by police were ultimately judged to be "unfounded" by the research team—potentially leading to a substantial inflation of aggregate crime statistics, as reported by police agencies.

A vast majority of the error found in this study involved the *undercounting* of Group A Hate Crimes (-67.35%).

Most of the error in hate crime reporting is from undercounting (i.e., crimes *not* originally identified as hate crimes/incidents by the police, and later determined by the reviewers to contain indicators of bias).

Focus group results support the notion that police officers adhere to various cognitive (mis)conceptions of bias to make decisions on the classification of crimes involving hate.

References

Bell, J. (2002). *Policing hatred: Law enforcement, civil rights, and hate crime*. New York: New York University Press.

Dewey, J. (1910|1997). *How we think*. Mineola, NY: Dover Publications, Inc.

Ericson, R. (1981). *Making crime: A study of detective work*. Toronto: Butterworths.

Federal Bureau of Investigation (1997). *Hate crime data collection guidelines*. Washington, D.C.: U.S. Government Printing Office.

McDevitt, J., Cronin, S., Balboni, J., Farrell, A., Nolan, J., and Weiss, J. (2003). *Bridging the information disconnect in bias crime reporting*. Washington, D.C.: Bureau of Justice Statistics, U.S. Department of Justice.

Nolan, J. and Akiyama, Y. (1999). An analysis of factors that affect law enforcement participation in hate crime reporting. *Journal of Contemporary Criminal Justice*, 15 (1): 111-127.

Nolan, J. and Akiyama Y. (2002). Assessing the climate for hate crime reporting in law enforcement organizations: A force-field analysis. *The Justice Professional* 15(2):87-103.

Nolan, J., Haas, S., Lester, T., Kirby, J., and Jira, C. (2006). *Establishing the "statistical accuracy"of Uniform Crime Reports (UCR) in West Virginia*. Charleston, WV: Criminal Justice Statistical Analysis Center, Division of Criminal Justice Services, Department of Military Affairs and Public Safety. Available online at www.djcs.wv.gov/SAC.

Nolan, J., Haas, S., and Napier, J. (2011). Estimating the impact of classification error on the "statistical accuracy" of Uniform Crime Reports. *Journal of Quantitative Criminology*. Available online: http://www.springerlink.com/content/08272p151rn78445/.

Nolan, J., McDevitt, J., Cronin, S., and Farrell, A. (2004). Learning to see hate crimes: A framework for understanding and clarifying ambiguities in bias crime classification. *Criminal Justice Studies* 17(1):91-105.

Shively, M., McDevitt, J., Cronin, S., and Balboni, J. (2001). *Understanding the prevalence and characteristics of bias crime in Massachusetts high schools, final report*. Boston, MA: Northeastern University, College of Criminal Justice.

Acknowledgments

This study would not have been possible without the cooperation and support of the local police agency's officials and staff. The authors greatly appreciate the time spent pulling records, answering questions, and participating in the focus group.

A special thanks is also extended to Simon Bauer-Leffler, Jessica S. Napier, and Monika Sterling with the Statistical Analysis Center and Susan Bennett, Stacia Gillard-Matthews, and Rachel Stein with West Virginia University for their invaluable assistance in reviewing records.

Funding Source

This project was supported by subaward agreement #WV12-2009-001, awarded by the Justice Research and Statistics Association (JRSA). This agreement is for the performance of a portion of the work originally awarded to JRSA from the Bureau of Justice Statistics, U.S. Department of Justice, award #2009-BJ-CX-K034. The views expressed in this report are those of the authors and do not necessarily reflect the opinions of the U.S. Department of Justice, the Bureau of Justice Statistics, or the West Virginia Division of Justice and Community Services.

Recommended Citation

Haas, S. M., Nolan, J. J., Turley, E., and Stump, J. (July 2011). *Assessing the validity of hate crime reporting: An analysis of NIBRS data*. Charleston, WV: Criminal Justice Statistical Analysis Center, Office of Research and Strategic Planning, Division of Justice and Community Services. Available online at www.djcs.wv.gov/SAC.

DJCS Administration

J. Norbert Federspiel, *Director*

Jeff Estep, *Deputy Director*

ORSP Administration

Stephen M. Haas, Ph.D., Director

1204 Kanawha Boulevard, East

Charleston, WV 25301

Phone: 304-558-8814

Fax: 304-558-0391

www.djcs.wv.gov