



The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Cases Associated with Violence in the National Missing and Unidentified Persons System (NamUs): The Examination of Circumstances & Characteristics Project

Author(s): Steven P. Hafner, Sc.D., B.J. Spamer, M.F.S., Bruce Budowle, Ph.D.

Document Number: 301932

Date Received: August 2021

Award Number: 2016-MU-BX-K007

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Cases Associated with Violence in the National Missing and Unidentified Persons System (NamUs)

The Examination of Circumstances & Characteristics Project

Submitted December 2020

University of North Texas Health Science Center
3500 Camp Bowie Blvd
Fort Worth, Texas 76107-2699

Authors

Steven P. Hafner, ScD, Research Scientist¹

B.J. Spamer, MFS, Executive Director, NamUs Operations²

Bruce Budowle, PhD, Professor and Executive Director, University of North Texas
Center for Human Identification³

This project was supported by Award No. 2016-MU-BX-K007, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this document are those of the authors and do not necessarily reflect those of the Department of Justice.

¹ Steven.Hafner@unthsc.edu

² BJ.Spamer@unthsc.edu

³ Bruce.Budowle@unthsc.edu

Table of Contents

Executive Summary.....	1
I. Introduction.....	3
II. Missing Persons.....	7
Background.....	7
Methods.....	8
Results.....	14
Recommendations.....	31
Conclusion.....	33
III. Unidentified Persons.....	34
Background.....	34
Methods.....	35
Results.....	41
Recommendations.....	52
Conclusion.....	54
IV. Special Section: Missing and Unidentified Indigenous Persons.....	55
Background.....	55
Methods.....	56
Results: MP.....	56
Results: UP.....	72
Recommendations.....	72
Conclusion.....	72
V. General Conclusion.....	73
VI. References.....	75
Appendix A - Technical Note.....	78
Appendix B - Final Classification Keywords.....	90

Executive Summary

This report details the approach, findings, and recommendations of a systems improvement and expansion project for a national public safety program – the National Missing and Unidentified Persons System (NamUs). Based on analyses of data from the NamUs database, recommendations are made that could be implemented to improve response to and resolution of cases within the database in which violence is a contributing factor.

The National Missing & Unidentified Persons System (NamUs)

NamUs is an asset of the U.S. Department of Justice and serves as a centralized database and resource center for cases of missing, unidentified, and unclaimed persons. NamUs offers a technological component – the database located at www.NamUs.gov – that can be used by law enforcement, medical examiners/coroners, allied professionals, and the public to manage, track, and resolve cases. Additionally, NamUs offers or provides funding support for various services, such as forensic services (DNA, fingerprints, forensic odontology), analytical services, and victim services to assist in the resolution of cases and to help family, friends, and other persons affiliated with a case in coping with stress faced by such persons in the wake of a missing, unidentified, and/or unclaimed persons event.

Violence Against Persons in NamUs

This report includes six sections, a technical appendix, and an appendix of classification keywords. Part I provides a general introduction to the NamUs program, an overview of violence and victimization in the United States, and a description of the nexus between NamUs cases and violence and victimization.

Part II of this report focuses on missing persons (MP) cases within the database. Data from 34,507 resolved (archived) and unresolved (published) cases were used to complete a qualitative text mining case classification process to classify cases as violent or non-violent using keywords indicative of violence (e.g., fight, hit). Then, wider case information was used to complete a descriptive analysis comparing violent and non-violent cases across various case characteristics. This analysis was stratified by gender of the case (male, female). Key findings and corollary recommendations are presented with nine recommendations arising from five key findings outlined within this section.

Part III of this report focuses on unidentified persons (UP) cases within the database. Data from 15,841 resolved (archived) and unresolved (published) cases were used to classify cases as violent or non-violent based both on manner of death (MOD) and qualitative text mining. All cases with a MOD as homicide listed were classified as violent, as well as additional cases uncovered through the text mining using keywords indicative of case violence. Then, broader case information was used to descriptively analyze the violent versus non-violent cases by decedent gender to identify any differences between violent and non-violent cases based on select case characteristics. Key results and associated recommendations are presented with nine recommendations arising from five key findings.

Part IV of this report is a dedicated special section to violence against American Indian and Alaska Native (AI/AN) women and men within the database. Similar procedures as used in parts II and III of the report were used to complete this section. Overall, there were only 792 AI/AN MP cases and 159 UP cases. Although MP results are presented, most results involve at

least one cell with a count of less than ten persons. Results for UP cases are not presented due to the lack of robust numbers of cases on which to make comparisons and conclusions. Thus, no recommendations are made for systems improvement and expansion within this section. Should more AI/AN cases populate the database, a more extensive analysis with attendant recommendations may be completed.

Finally, Part V of this report provides a general conclusion to the work presented. Additionally, important limitations to the report overall are outlined for consideration. Part VI presents references used throughout the report. The appendices to this report provide greater context and information on the design and implementation of this project. Appendix A presents a technical note that provides more detail as to the classification and analysis processes for both MP and UP cases. Appendix B provides the keyword classification dictionaries used for the classification section of the respective sections.

I. Introduction

The National Missing & Unidentified Persons System

The National Missing and Unidentified Persons System (NamUs) is a U.S. Department of Justice asset funded and administered by the National Institute of Justice (NIJ). NIJ is one of two science agencies housed within the Office of Justice Programs. NamUs is managed through a cooperative agreement with the University of North Texas (UNT) Health Science Center's UNT Center for Human Identification (UNTCHI). NamUs brings together law enforcement, medical examiners, coroners, forensic experts, families, and the public to help resolve missing, unidentified, and unclaimed person cases throughout the nation – what has been called “the nation’s silent mass disaster” (Ritter, 2007). The NamUs database provides criminal justice users a secure, online system to store, share, and compare sensitive case information, and public access allows family members and other public stakeholders a mechanism to report cases and participate in the search for potential leads. As of October 25, 2020, the NamUs database contained 19,501 published missing person (MP) cases, 13,486 published unidentified person (UP) cases, and 10,011 cases of unclaimed decedents who have been identified by name, but whose next of kin have not been located for death notification. A published case is a case that has been entered by the public, law enforcement, a medicolegal authority, or some other affiliated personnel and that has been vetted as a legitimate case that can be made publicly available on the NamUs website.

Case Entry and Search

Missing person cases are entered into NamUs by both public and professional (i.e., law enforcement) users, while UP cases are entered only by medical examiners/coroners (ME/Cs), or their designees. Before being published for public viewing and searching, MP cases entered into NamUs must be reported to law enforcement and that investigating agency must validate and give NamUs permission to publish the case. Although there is an unclaimed persons database that is part of NamUs, it was not used for this project.

All active cases that have been vetted (e.g., have an active MP report) and published can be searched at www.NamUs.gov by any registered user or non-registered visitor. Information displayed to the public is limited, as there are data fields that are restricted from public view if deemed sensitive to a case investigation or to a MP. The public search component of NamUs is a defining feature of the database that allows for stakeholders from multiple fields to assist in the resolution of cases.

NamUs Services

As a national resource center for MP and UP cases, NamUs provides forensic services to assist with case resolutions, including forensic odontology and fingerprint examination, and leverages DNA analyses and anthropological services performed by subject matter experts within the UNTCHI forensic laboratories. A team of Regional Program Specialists (RPSs) serves as a force multiplier for criminal justice agencies across the country, providing investigative support and guidance to assist with case investigations. RPSs also directly support families of missing persons by connecting families to investigating agencies to file missing person reports, facilitating reference sample collections for DNA analyses, and participating in missing person events across the country to intake new cases and augment existing NamUs case files. Finally,

the NamUs Victim Services Division addresses the needs of family, friends, and others impacted by the death or disappearance of a loved one by facilitating trauma-informed therapies, a peer support network, reunifications, and referrals for other services.

NamUs Impact

By connecting people, information, forensic science, and technology, NamUs supports families and provides medical examiners, coroners, and criminal justice professionals with access to the investigative and scientific tools needed to find missing persons, help identify decedents and victims of crime, assist in resolving criminal cases, and reduce violent crime and human trafficking.

A Nexus with Violence and Victimization

Violence and Victimization

Globally, violence is a persistent phenomenon that affects both women, men, and children as seen in both public health and criminal justice data. According to the 2010 National Intimate Partner and Sexual Violence Survey (NISVS), 9.4% of women and 2.2% of men had been raped by an intimate partner in their lifetime while 16.9% of women and 8.0% of men have experienced sexual violence other than rape by an intimate partner. 24.3% of women and 13.8% of men have experienced severe physical violence by an intimate partner, 10.7% of women and 2.1% of men have been stalked by an intimate partner, and 48.4% of women and 48.8% of men have experienced psychological aggression by an intimate partner (Breiding, Chen, & Black, 2014). Using 2011 data from NISVS, Breiding and colleagues (2014) found that 19.3% of women and 1.7% of men have been raped, 43.9% of women and 23.4% of men have experienced other forms of sexual violence, 15.2% of women and 5.7% of men have experienced stalking, and 22.3% of women and 14.0% of men have experienced severe physical violence by an intimate partner in their lifetime. Finally, and most recently, the 2015 NISVS found that 21.3% of women and 2.6% of men experienced completed or attempted forced penetration in their lifetime, and 16.0% of females and 5.8% of men experienced stalking at some point in their lifetime. Regarding intimate partner violence, 18.3% of women and 8.2% of men experienced contact sexual violence by an intimate partner, 30.6% of women and 31.0% of men experienced physical violence by an intimate partner, 10.4% of women and 2.2% of men experienced stalking by an intimate partner, and 36.4% of women and 34.2% of men experienced psychological aggression by an intimate partner in their lifetime (Smith et al., 2018)

Additionally, according to the Centers for Disease Control and Prevention's (CDC) Web-based Injury Statistics Query and Reporting System (WISQARS), in 2018, homicide was the fourth leading cause of death for females ages one to four and five to nine, the fifth leading cause of death for females ages 10-14 and 25-34, the third leading cause of death for females 15-24, and the eighth leading cause of death for females 35-44. For males, homicide was the second leading cause of death for ages one to four, the fourth leading cause of death for ages five to nine and 10-14, the third leading cause of death for ages 15-24 and 25-34, the fifth leading cause of death for ages 35-44, and the ninth leading cause of death for ages 45-54 (National Center for Injury Prevention and Control, 2020). Furthermore, for women, as many as 38% of murders of women are committed by an intimate partner globally (World Health Organization, 2016). In cases where the victim's relationship to the offender was known, 93% of murdered women were killed by a male they knew, with 63% of those victims being the wife or intimate partner of the

perpetrator (Violence Policy Center, 2016). Additionally, according to the Federal Bureau of Investigation's (FBI) National Incident-based Reporting System, males were far more likely to be perpetrators of violent crime (79%) and more likely to be victims of violent crime (51%) (Crime Data Explorer, 2019). Similarly, the FBI's Supplemental Homicide Report for 2018 shows 12,495 males were victims of homicide compared to 3,686 females, and that the gender of the oldest offender was more often male (n = 12,034) than female (n = 1,711) (Puzzanchera, Chamberlin, & Kang, 2020).

Capturing Violence in NamUs Cases

For MP cases, NamUs does include built-in functionality for case entrants to indicate if some type of nefarious activity was related to the case. This functionality is operationalized through the inclusion of a "foul play" indicator that may be checked by the case entrant, NamUs staff, or investigative personnel (if linked as a Case Contributor) if it is believed that violence or coercion may have played a role in the missing persons case. For UP cases, foul play would be represented by including a MOD as "homicide" by the case entrant. It is important to note, however, that there may be wide variation in the use of the "foul play" indicator depending on if a case is an ongoing investigation, depending on the particular case entrant, who may always or never check the "foul play" box, or for other reasons related to case investigation.

As of October 25, 2020, of the unidentified decedent cases reported to NamUs, 2,691 of the 18,018 active and archived cases were reported to be victims of homicide. Another 7,116 of all published and archived cases were reported to have an "undetermined" MOD at the time of entry into NamUs; it is possible that many of these cases may contain circumstances or cause of death information which suggest a homicidal MOD is probable. Additionally, of the 42,432 active and archived MP cases in NamUs, 3,106 indicated that foul play was involved, and 776 indicated that foul play was possible.

Phase 1: Examination of Circumstances and Characteristics (ECC) Project

Purpose of the Project

There is a clear overlap with the serious problem of violence with a portion of the cases in NamUs. Although NamUs currently focuses on all MP and UP cases, there are good reasons to address the relationship and impact of violence within the subject populations. First, given its current mission and functionalities, NamUs potentially could expand its role to provide services, information, and guidance to men and women and their families who have been subjected to violence. Additionally, information gleaned from this project could be used by personnel and agencies to assist in the investigation and resolution of cases involving violence by providing key context and information specific to violent cases that could support an investigation.

To improve and expand the NamUs program, including its services, information, and guidance to persons of all ages and their families, a clear understanding of case characteristics specific to missing and unidentified cases is critical. Through comparison with cases involving either men or women, such characteristics may be identified and may be used to create targeted outreach and programmatic activities for women and men, including their families. Given that NamUs includes information on both women and men, it is uniquely situated as a data source to provide the level and type of comparisons necessary for comprehensive and in-depth investigation that will yield results that potentially can be put into practice.

Goals of the Project

Although there is a nexus between cases in NamUs and the issues of violence and victimization, there has been no comprehensive assessment 1) to determine the extent to which cases in the NamUs system may involve violence or how such violence may differ by gender, 2) to ascertain case characteristics (e.g., perpetrator identified, victim-perpetrator relationship, etc.) for both MP and UP cases, 3) to determine criminal justice involvement (e.g., arrest, prosecution) of identified perpetrators, and 4) to make recommendations for system improvement, enhanced data collection, and specific services for MP and UP cases involving violence. Because of the outreach that NamUs has, there is a need to assess the current role the program serves with violence against persons, and more importantly, develop and enhance NamUs to better serve the victims of violence against persons.

Specific Aims

To achieve its overarching goals, this phase of the ECC project has two specific aims:

1. To classify MP and UP cases as violent or non-violent through the development and implementation of a qualitative, automated case classification method.
2. To determine if differences, including gender-specific differences, between violent and non-violent cases exist across case characteristics.

Project Approach and Outcomes

To achieve the aims of this project, a novel strategy using text mining and case classification was used. These methods were used for MP and UP cases separately and did not include unclaimed persons. In the first phase of this project, the project sought to ascertain the extent to which violence was present in active and archived cases within the NamUs system, including the category of violence and how violent cases may differ across characteristics of the case (e.g., demographics, etc.). This effort was undertaken in order to provide a more comprehensive picture of the extent of violence, types of violence, unique needs that may impact victims or perpetrators using violence, and how they may impact system and service improvements\.

Phase 2 of this project will ascertain criminal justice involvement within violent cases, as well as make recommendations for systems improvement. Given the aims and methods of this phase of the project are related to systems improvement and expansion, it was deemed Not Human Subjects Research by the North Texas Regional Institutional Review Board.

In fulfilling these aims, UNTCHI will be able to make recommendations for systems improvement and expansion to include enhanced and more relevant data collection and services that could benefit the NamUs program as it intersects with case violence. Improvements may include new or expanded data fields for more comprehensive data collection and recommendations to the various service divisions (e.g., victim services) to improve service provision to violent cases, which may differ from the requirements of non-violent cases. Overall, this project will benefit the NamUs program, its users, and the MP and UP who populate the system through enhanced understanding of how violence plays a role in case data collection, management, and resolution.

II. Missing Persons

Background

Each year, over 600,000 persons are reported missing in the United States (FBI, 2020). As of October 25, 2020, there were 19,501 active MP cases within the NamUs database. Fortunately, most MPs are recovered alive in a short period of time, a fact evidenced by MP statistics from the FBI's National Crime Information Center for 2019. These statistics show that there were roughly 609,000 MP entries into the system with 87,438 cases active as of December 31, 2019. Over the course of the year, 607,104 cases were purged from the system due to the MP being located, the MP voluntarily returning home, or the record being found invalid (FBI, 2020). Thus, long-term (i.e., one year or longer) MP events occur infrequently as a proportion of the total number of MP cases filed each year; however, it remains unclear as to the extent to which violence contributes to case duration and clearance.

While many cases represent voluntary disappearances, misunderstandings regarding whereabouts, miscommunications, or some other benign factor that has led a person or people to believe another person is missing involuntarily, some cases do involve violence victimization or perpetration. Bonny and colleagues (YEAR) classified MP events into more formal categories reflecting these types of events. In their classification, MPs are considered missing due to dysfunction (e.g., mental health issue), to escape (e.g., leave due to violent relationship), or are unintentionally missing (e.g., missing under the influence of a third party). In an Australian study, Henderson and colleagues (2000) outlined four possible reasons for a MP event: independence or rebellion (e.g., conflict over parental authority), safety concerns (e.g., abduction, accident), unintentional (e.g., miscommunication), and escaping adverse consequence (e.g., leaving a violent environment). In each typology, a specific category for cases that may involve violence experienced or perpetrated by the MP is given, indicating that such cases are prevalent and contribute to the overall burden of MP cases.

Violence and Missing Persons

Although it is not entirely clear how many MP cases involve foul play, as seen in the typologies of missing persons, there are instances in which foul play or violence may be involved. As of October 25, 2020, 2,483 of the 19,501 active NamUs MP cases (12.7%) indicated that foul play may have been a factor in the person's disappearance. However, the true extent to which violence or nefarious circumstances contribute to MP cases as a whole is little understood. Additionally, the types of violence that may contribute to MP cases is understudied. While it may be believed that physical violence is most common in such instances, other forms of violence, such as psychological aggression (e.g., arguments), may also contribute to the disappearances in particular cases of voluntary MP events.

The current project sought to better understand the burden of violence in MP cases within a national database. To do so, it relied on case file information from active and archived cases within a database. Violence was defined broadly to include both physical and non-physical aggression that may result in physical, psychological, emotional, or other harm to an individual. In completing this study, a broader, more nuanced picture as to the extent of violence within cases within the database was elicited, as well as a better understanding as to how violence may be related to other characteristics of a case.

Methods

This section describes briefly the methods used to complete case classification and case analysis. A more detailed treatment of methods, including more nuanced information on dictionary building, misspellings, coding, and reporting is provided in Appendix A – Technical Note.

Project Data

Data for this project came from the MP sections of the NamUs database. Data include all published and archived cases as of the pull date of August 29, 2019. Data were pulled by the Senior Business Intelligence Developer and delivered to the Project Director using the secure, password-protected sharing service OneDrive. Data were delivered as a single Microsoft Excel file with three separate sheets. These sheets included Case Information that provided demographic and case details on the missing person; Circumstances that provided text entries for general circumstance information available for public view, Agency Notes available only for professional users, and Circumstances of Resolution that provided information on the outcome of a resolved case; and Agency Information that provided details as to the entering and contributing parties on the case. These three sheets were saved as separate Microsoft Excel files. Data were maintained on an encrypted, password-protected, UNTHSC-provided computer.

Case Classification

Initial Classification Dictionary

To classify cases as violent or non-violent, a classification dictionary was created in WordStat 8.0.20 (Provalis Research; Montreal, Quebec, Canada) that included keywords relevant to MP cases. See Appendix B – Final Classification Keywords for a list of keywords used to classify cases as violent. This dictionary was developed collaboratively by the project team in order to ensure the best coverage and most exhaustive list of potential keywords that were relevant to the classification of cases as violent or non-violent. The dictionary for this project was also built using several sources, including keywords identified by project staff, through thesauri both online and within the software, and from various external sources, including the CDC's National Intimate Partner and Sexual Violence Survey, the Conflict Tactics Scale-2 (Straus et al., 1996), and CDC's Youth Risk Factor Behavioral Surveillance System and Behavioral Risk Factor Surveillance System. In total, 100 words were initially identified and included in the dictionary under six categories: general violence (40 words), psychological aggression (13 words), physical violence (30 words), coercive control (7 words), sexual violence (4 words), and stalking (6 words). Amendments to the initial dictionary were based on classification outcomes. The final dictionary shown in Appendix B includes 124 words that were identified and included in the final dictionary under six categories: general violence (46 words), psychological aggression (17 words), physical violence (40 words), coercive control (10 words), sexual violence (5 words), and stalking (6 words).

Classification Runs

In total, there were three separate classification runs for MP cases. The order of steps and overall approach to these runs varied slightly and were based on lessons learned from the previous runs.

General Classification Procedures. MP case classification involved three runs of the data following the same general procedures and was carried out using WordStat. More detailed information on each classification run can be found in Appendix A. First, data were delivered

and assembled to include all relevant circumstantial text fields – Circumstances Text, Agency Notes, and Circumstances of Resolution Text. Then, data were cleaned by removing any cases that did not have any information in any of the circumstance fields, that only indicated some form of “unknown,” or that contained only numbers or symbols. Cleaned data were then imported into WordStat for classification. Several classifications were tested, including with and without an exclusion list based on lessons learned from previous runs of the data and included keywords that created false positives, which did not indicate true violence. In Run 2, a misspellings check was added to the procedures. This check was completed using WordStat’s built-in functionality and provided a list of potential misspellings. Each potential misspelling was checked using a “keyword-in-context” feature to ensure the misspelling truly represented a misspelled word, a proper noun, or some otherwise correct spelling for a different word. The final run of the data included use of the exclusions list previously created, as well as corrected misspelled words.

Following the final runs, data were exported from WordStat and imported into Stata. A random sample of 10% was drawn for second coding by a human coder (Hafner). Percent agreement was calculated between WordStat and the human coder. Following the calculation of percent agreement, a disagreements check was performed to understand discrepancies in coding between WordStat and the human coder. Through this process, and through review of the dictionary by co-investigator (Spamer), it was determined that there were additional words that needed to be added to the dictionary. As such, it was determined that second and third runs of the data would be necessary. In the third run of the data, computer-human agreement was found to be 88%. This level of agreement was found to be close to high, which Schweta, Bajpai, and Chaturvedi (2015) define as a percent agreement of 90% or above.

In the third run, tertiary coding was completed by a second human coder (co-investigator Spamer) using a subsample of 20% of the 10% of cases originally sampled for second coding. Percent agreement was then calculated between the two human coders and was found to be 93%. This level of agreement is considered to be high (Schweta, Bajpai, & Chaturvedi, 2015).

Final Violence Coding. Following tertiary coding, given the high levels of agreement between WordStat and the human coder, and the high level of agreement between the two human coders, a final coding of all cases classified as violent by WordStat was performed. In doing so, only cases classified as violent by WordStat were examined to ensure they truly represented violence and that the classification as violent was not spurious based on keywords being used in a different context (e.g., DNA “hit”). Overall, 7,659 individual cases classified as violent were manually read with 3,493 cases being retained as truly violent cases. Later, 86 cases ruled as homicides by ME/Cs but not classified as violent during the classification process and based on circumstantial information were manually added to the violent case count, bringing the total number of violent cases to 3,579.

Case Analysis

Following completion of case classification, the data were prepared for analysis such that meaningful comparisons could be made between violent and non-violent cases for both female and male cases.

Final Dataset Creation

To arrive at a final analytic dataset, all non-relevant data fields pertaining to the final coding moving forward were dropped. This action created final coded data that were imported into

Stata. In Stata, further data fields were removed that were not pertinent for the case analysis (e.g., CircumstancesText), leaving a set of numerical indicator variables denoting the presence of violence. Given that cases could have more than one line per case of circumstantial information (e.g., Agency Notes), these numerical values had to be collapsed and summed to arrive at one line per case in the final dataset. Following this collapse, indicator variables were created and dichotomized, and data were checked for consistency. No immediate errors were uncovered. The resulting dataset was then merged on Case ID with the Case Information dataset saved initially when the data were received. This final dataset contained 34,410 cases; however, one case did not have any associated Case ID. This case was dropped resulting in 34,509 cases. Further, given that this analysis intended to look at results by gender, cases with no gender indicated (n = 1) or with “unsure” gender (n = 1) were also dropped. These steps resulted in a final analytic dataset containing 34,507 unique cases.

Variables

Variables used in the analysis stage of this project included violence variables, variables for all cases, variables for resolved cases only, and variables for unresolved cases only. Each set of variables is described below.

Violence Variables. Violence variables (Table 1) included one indicator variable denoting if violence of any type was present within the case and six different types of violence.

Table 1. Violence Variables

Variable	Definition	Coding	N
Finalviolence	Indicates whether or not any type of violence was present in a case	0 = No 1 = Yes	30,928 3,579 Total: 34,507
GV (GENERAL_VIOLENCE)	Includes forms of violence that could not definitively be classified within other categories, such as physical or verbal violence. This type of violence includes terms such as “anger,” “altercation,” “dispute,” and “aggression”	0 = No 1 = Yes	32,840 1,667 Total: 34,507
PV (PHYSICAL_VIOLENCE)	Includes forms of violence that cause bodily injury through person-to-person contact or through the use of a weapon. This type of violence includes keywords such as “hit,” “slap,” “injure,” “stab,” and “shoot”	0 = No 1 = Yes	32,045 2,462 Total 34,507
SV (SEXUAL_VIOLENCE)	Includes violence that encompasses forced or coerced unwanted sexual activity. This type of violence includes terms such as “rape,” “molest,” and “sexual assault”	0 = No 1 = Yes	34,474 33 Total: 34,507
ST (STALKING)	Includes violence that encompasses threatening, harassing, and unwanted behaviors that causes a person to have concerns or fear for safety. This type of violence includes terms such as “follow,” “spy,” and “unwanted communication”	0 = No 1 = Yes	34,345 162 Total: 34,507
CC (COERCIVE_CONTROL)	Includes violence that controls or threatens an individual’s actions, behaviors, or thoughts. This type	0 = No 1 = Yes	34,252 255

	of violence includes keywords “force,” “manipulate,” and “trafficking”		Total: 34,507
PA (PSYCHOLOGICAL_AGGRESSION)	Includes violence aimed at harming a person’s mental and emotional wellbeing, and includes terms such as “name call,” “humiliate,” “argue,” and “verbal disagreement”	0 = No 1 = Yes	34,018 489 Total: 34,507

All Case Variables. All case variables (Table 2) include information related to case demographics and general case characteristics.

Table 2. All Case Variables

Variable	Definition	Coding	N
Age_of_Case_groups2	Indicates period of time between date of last contact and present day	0 = <90 days 1 = 3-12 mo. 2 = 1-2 years 3 = 2-5 years 4 = 5-10 years 5 = 10-20 years 6 = 20+ years 7 = Unknown	339 1,715 2,535 6,941 9,590 6,180 7,194 13 Total: 34,507
StateName2	State where person in case went missing from (not necessarily state where missing person was found)	1 = Alabama 2 = Alaska 3 = Arizona 4 = Arkansas 5 = California 6 = Colorado 7 = Connecticut 8 = Delaware 9 = DC 10 = Florida 11 = Georgia 12 = Guam 13 = Hawaii 14 = Idaho 15 = Illinois 16 = Indiana 17 = Iowa 18 = Kansas 19 = Kentucky 20 = Louisiana 21 = Maine 22 = Maryland 23 = Massachusetts 24 = Michigan 25 = Minnesota 26 = Mississippi 27 = Missouri 28 = Montana 29 = Nebraska 30 = Nevada 31 = New Hampshire	453 456 1,568 529 4,710 1,370 802 93 107 2,650 712 3 182 304 846 404 183 220 504 492 175 383 378 1,719 322 199 737 139 217 495 63

		32 = New Jersey	761
		33 = New Mexico	372
		34 = New York	1,204
		35 = North Carolina	676
		36 = North Dakota	86
		37 = Northern Mariana Islands	2
		38 = Ohio	827
		39 = Oklahoma	809
		40 = Oregon	929
		41 = Pennsylvania	835
		42 = Puerto Rico	39
		43 = Rhode Island	41
		44 = South Carolina	406
		45 = South Dakota	76
		46 = Tennessee	887
		47 = Texas	2,380
		48 = Utah	281
		49 = Vermont	71
		50 = Virgin Islands	22
		51 = Virginia	550
		52 = Washington	1,249
		53 = West Virginia	178
		54 = Wisconsin	315
		55 = Wyoming	96
		Total:	34,507
FoulPlay2	Marker used to indicate if nefarious, violent, or otherwise potentially criminal activity led to the disappearance of a missing person	0 = No 1 = Possibly 2 = Uncertain 3 = Yes	3,932 309 6,589 2,544
		Total:	13,374
SexName2	Indicates biological gender of missing person as defined by reporting party	0 = Female 1 = Male	15,119 19,388
		Total:	34,507
RaceEthnicity2	Indicates race/ethnicity of missing person as defined by reporting party	0 = White / Caucasian 1 = American Indian / Alaska Native 2 = Asian 3 = Black / African American 4 = Hawaiian / Pacific Islander 5 = Hispanic / Latino 6 = Other 7 = Uncertain 8 = 2+ Races	20,349 713 764 6,013 20 3,478 364 117 2,685
		Total:	34,503
Age_Missing_Group2	Indicates age of missing person at time of missing event as defined by reporting party	0 = <18 1 = 18-20 2 = 21-30 3 = 31-40 4 = 41-50 5 = 50+ 6 = Unknown	9,314 2,020 7,157 5,519 4,165 5,425 907
		Total:	34,507

Resolved Cases Variables. Variables for resolved cases are shown in Table 3 and include information related to the case resolution information.

Table 3. Resolved Case Variables (N = 18,051)

Variable	Definition	Coding	N
CaseResolutionStatus2	Indicates if the final disposition of resolved cases was with a deceased or alive person	0 = Deceased 1 = Alive	4,563 13,473 Total: 18,036
NamUsAssisted	Indicator defined by professional user at case close-out to denote if NamUs was useful or an important factor in case resolution	0 = No 1 = Yes	15,593 2,357 Total: 17,950
MannerOfDeath2	Indicates manner of death for resolved cases in which the missing person was found deceased	1 = Accident 2 = Homicide 3 = Natural 4 = Pending 5 = Suicide 6 = Undetermined	55 124 25 113 59 232 Total: 608

Unresolved Cases Variables. For unresolved cases, variables included one variable related to the presence or absence of DNA samples uploaded and completed as shown in Table 4.

Table 4. Unresolved Case Variables (N = 16,456)

Variable	Definition	Coding	N
DNA_Indicator2	Indicates if at least one DNA sample analysis has been completed	0 = No 1 = Yes	7,502 8,954 Total: 16,456

Case Analysis

First, an indicator variable was created to show if the case was resolved or unresolved. This variable was created using the lifecycle status of the case – whether the case was archived (resolved) or published (unresolved) at the time the data were drawn. Subsequent analysis of the data for variables for resolved and unresolved cases was completed using only those variables within each respective classification. Then, univariate descriptive statistics for each variable were completed followed by univariate statistics stratified by MP gender. Finally, gender stratified (male, female) bivariate statistics comparing final case violence (yes, no) across levels of each variable were run in order to assess differences in each variable by the presence or absence of case violence for each case. All cases that had information on the variable of interest were included in the analysis. As such, the resulting tables do not display that same number of observations. All analyses were carried out in Stata 14.2 (StataCorp.; College Station, TX).

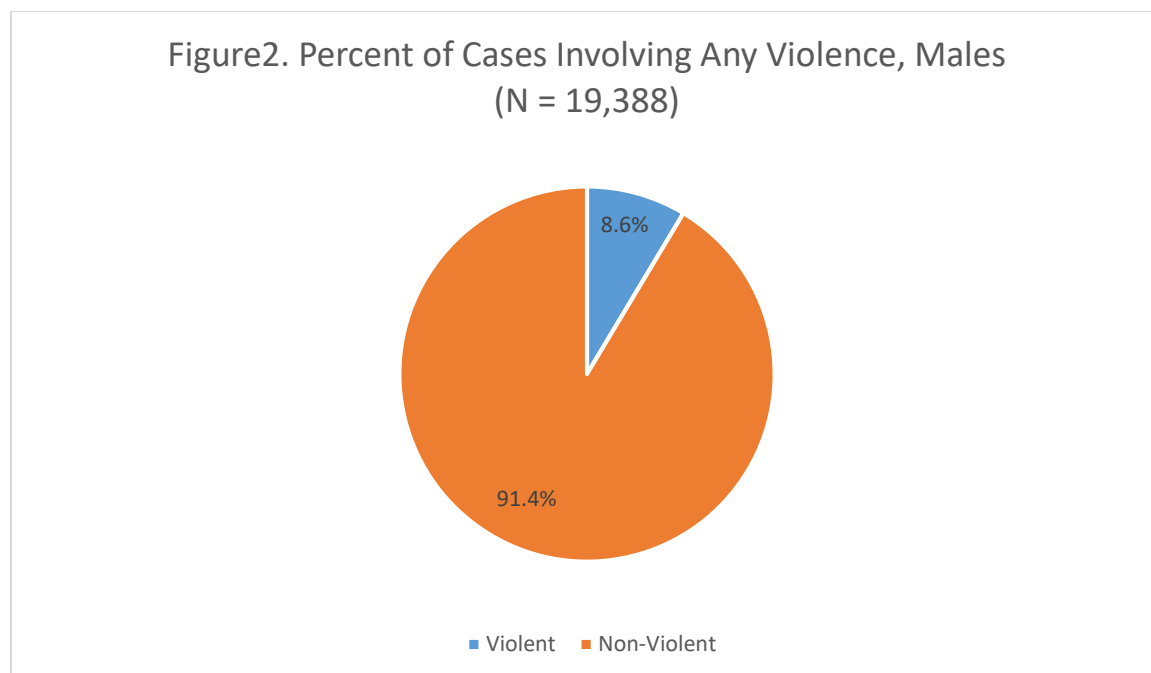
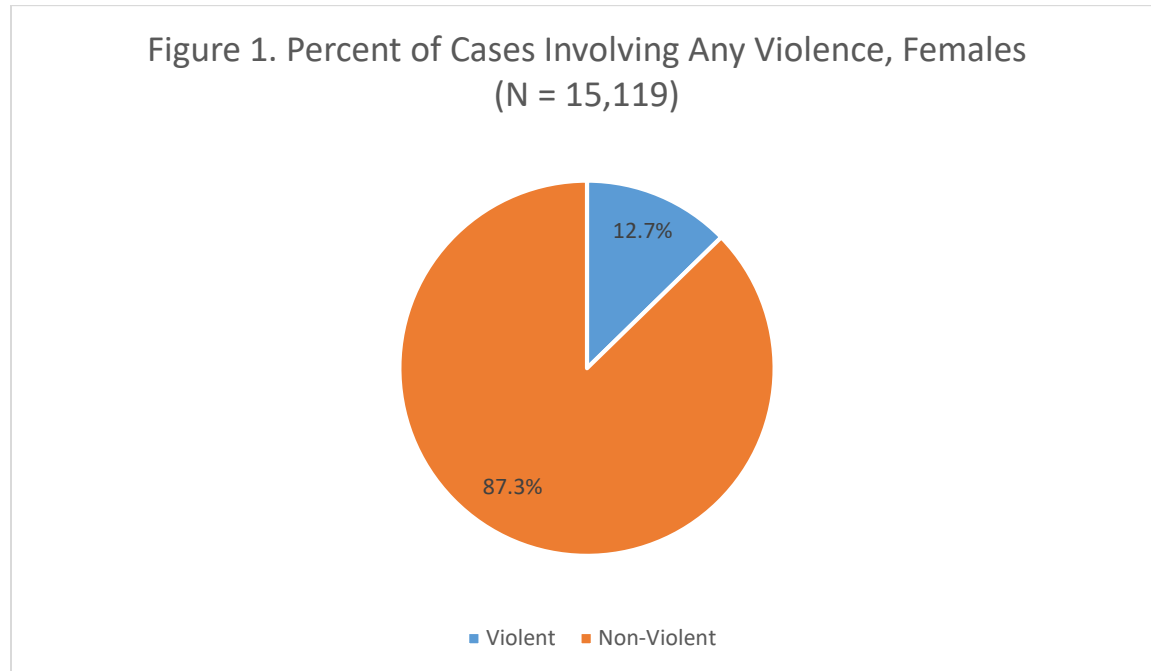
Results

Overall, the data analysis included 34,507 cases. Demographically, 15,119 (43.8%) were female, and 19,388 (56.2%) were male. White/Caucasian individuals made up the majority of cases by race/ethnicity representing 59.0% of all cases. The majority of cases were individuals under the age of 18 (27.0%). Tables 1-32 represent gender stratified results for victimization, as

well as gender stratified bivariate comparisons between final case violence and case characteristics by case resolution status. Results marked with an asterisk (*) include at least one cell count of fewer than 10 cases.

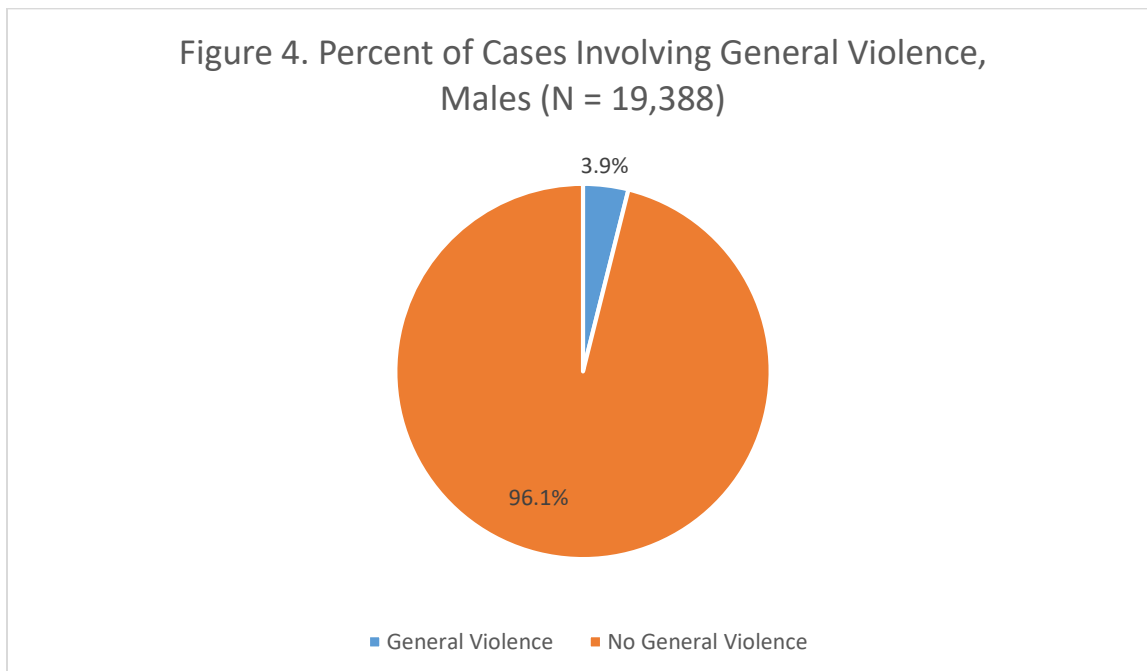
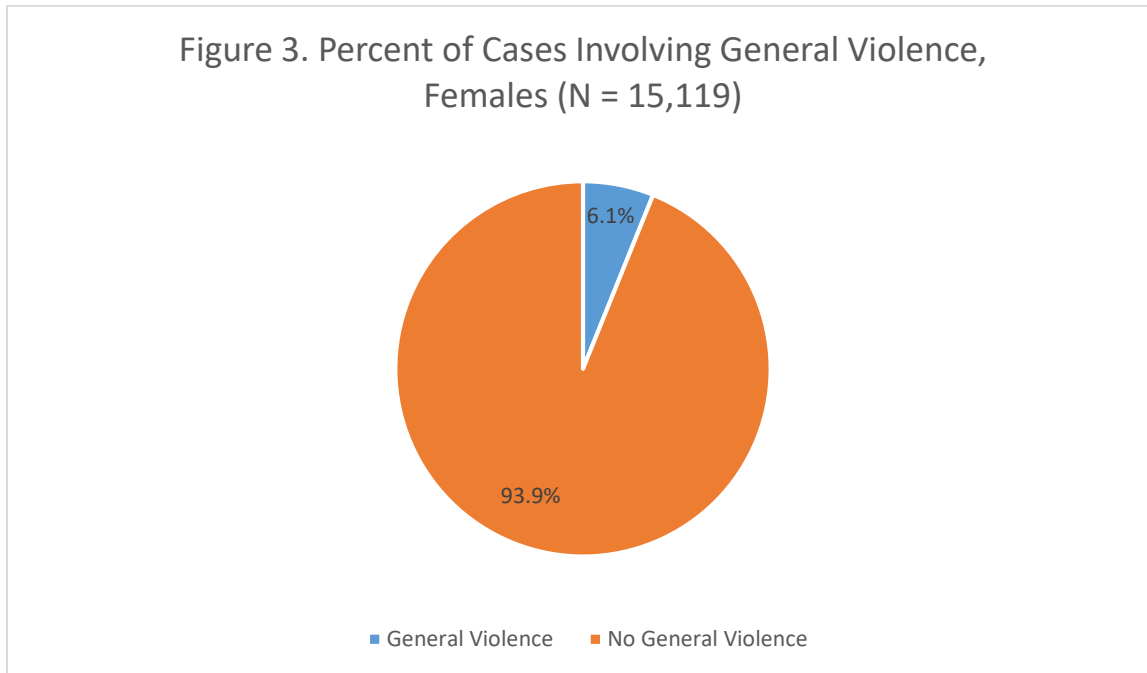
Overall Violence (N =34,507)

The percentage of females that experienced any form of violence was 12.7% (n = 1,916), while it was 8.6% (n = 1,663) for males (Figures 1 and 2). Thus, a total of 3,579 persons experienced violence, or 10.4% of all cases.



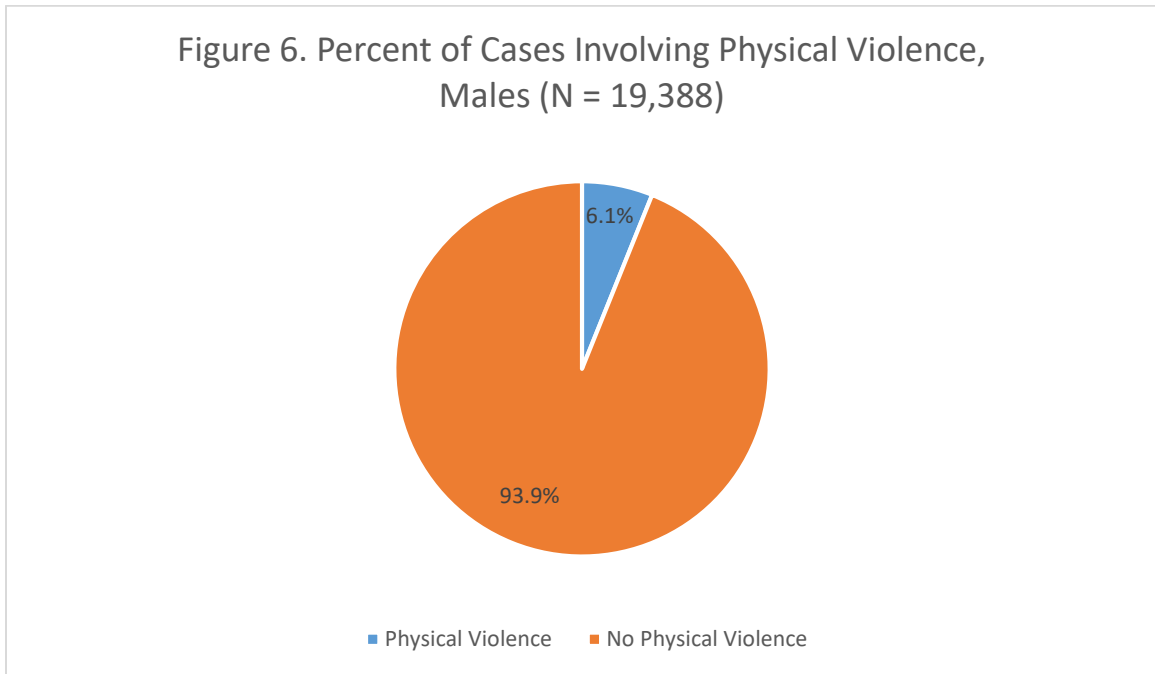
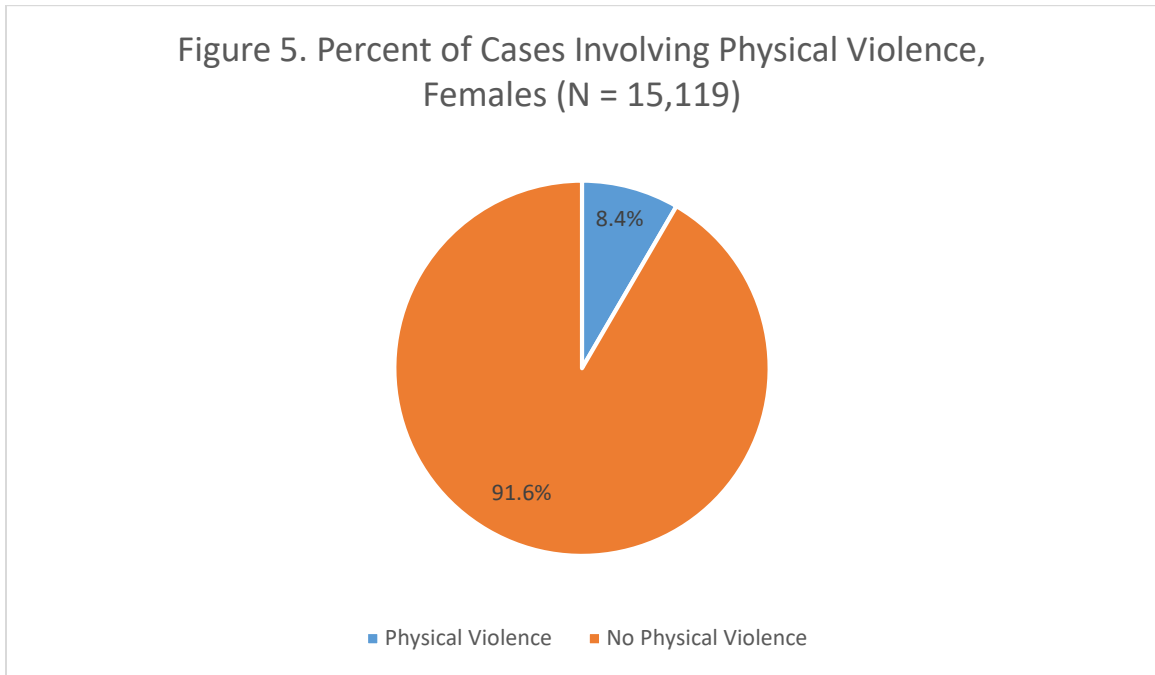
General Violence (N = 34,507)

As with overall violence, a higher proportion of females (6.1%, n = 919) than males (3.9%, n = 748) experienced general violence (Figures 3 and 40).



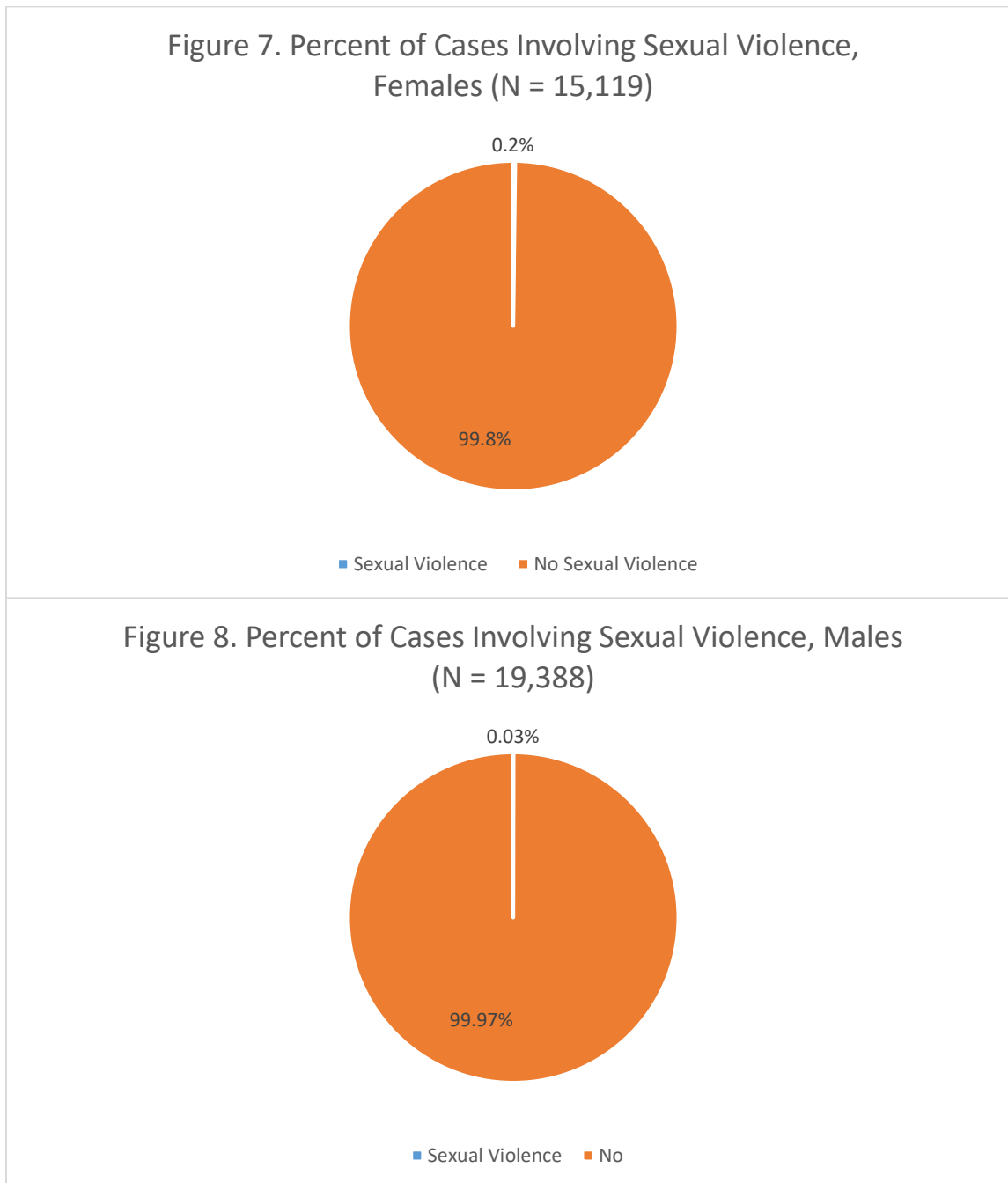
Physical Violence (N = 34,507)

Among females, 8.4% (1,275) experienced physical violence, while 6.1% (n = 1,187) of males experienced physical violence (Figures 5 and 6).



Sexual Violence* (N = 34,507)

There were 0.2% (n = 27) females who experienced sexual violence (Figure 7), while 0.03% (n = 6) of males experienced sexual violence (Figure 8).



Stalking Behavior (N = 34,507)

Among females, 0.6% (n = 95) experienced some type of stalking behavior (Figure 9), while 0.4% (n = 67) of males experienced stalking behavior (Figure 10).

Figure 9. Percent of Cases Involving Stalking Behavior, Females (N = 15,119)

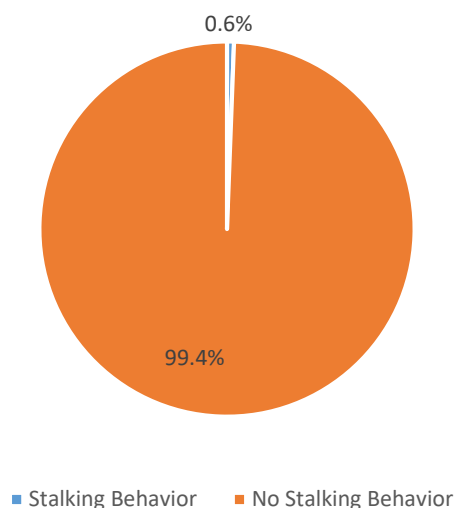
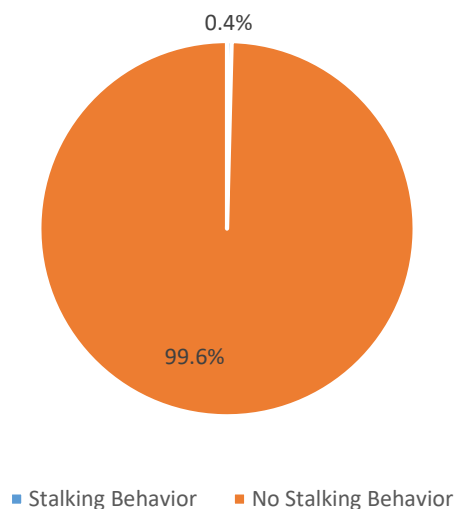


Figure 10. Percent of Cases Involving Stalking Behavior, Males (N = 19,388)



Coercive Control (N = 34,507)

Among females, 1.4% (n = 204) experienced coercive control (Figure 11), while 0.3% (n = 51) of males did as well (Figure 12).

Figure 11. Percent of Cases Involving Coercive Control, Females (N = 15,119)

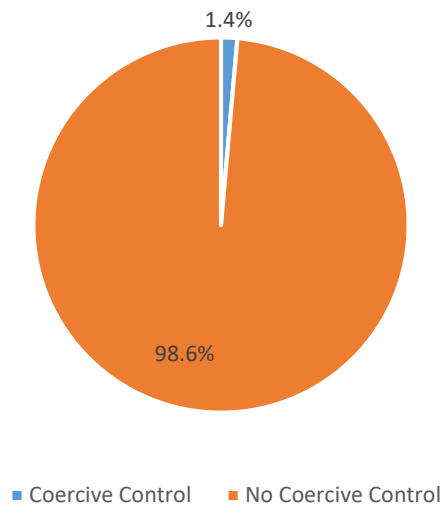
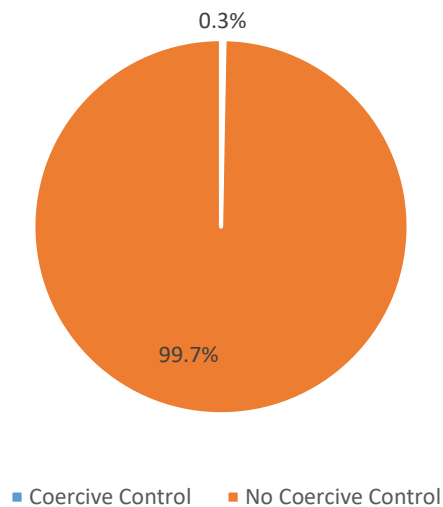


Figure 12. Percent of Cases Involving Coercive Control, Males (N = 19,388)



Psychological Aggression (N = 34,507)

For females, 1.8% (n = 269) experienced psychological aggression (Figure 1). For males, the proportion was lower at 1.1% (n = 220) (Figure 14).

Figure13. Percent of Cases Involving Psychological Aggression, Females (N = 15,119)

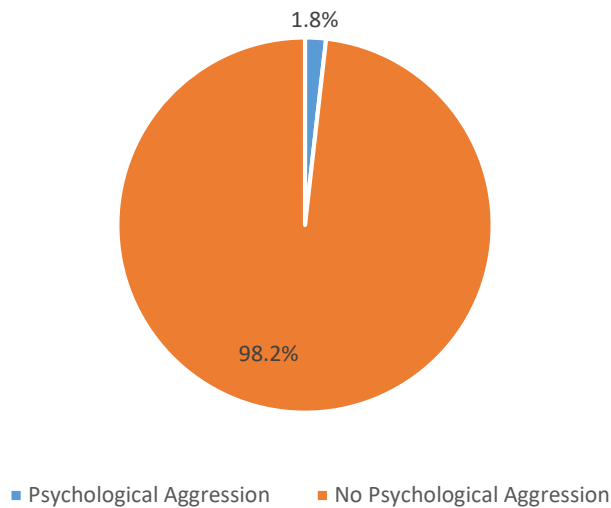
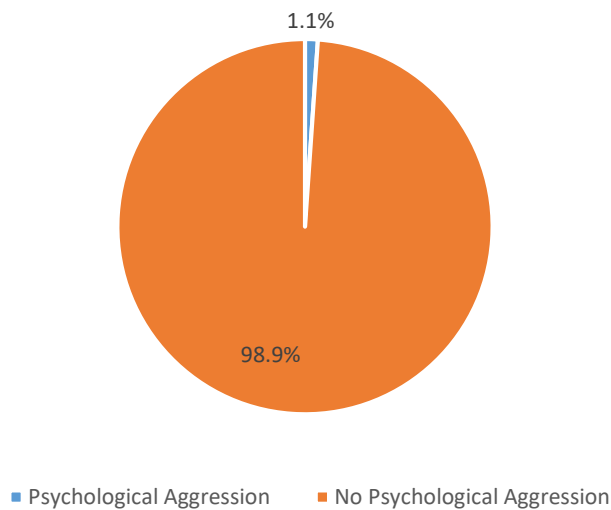


Figure 14. Percent of Cases Involving Psychological Aggression, Males (N = 19,388)



Age of Case (Groups) (N = 34,507)*

Among female non-violent cases, the highest proportion of cases were 5-10 years old (29.8%, n = 3,936) while the highest proportion of violent cases were 20+ years old (41.2%, n = 790) (Figure 15). This trend was mirrored among male violent cases where 33.8% (n = 562) were 20+ years old, while among male non-violent cases, the highest proportion of cases were observed among 5-10 year old cases (28.1%, n = 4,971) (Figure 16).

Figure 15. Comparison of Case Violence Across Age of Case Groups, Females (N = 15,119)

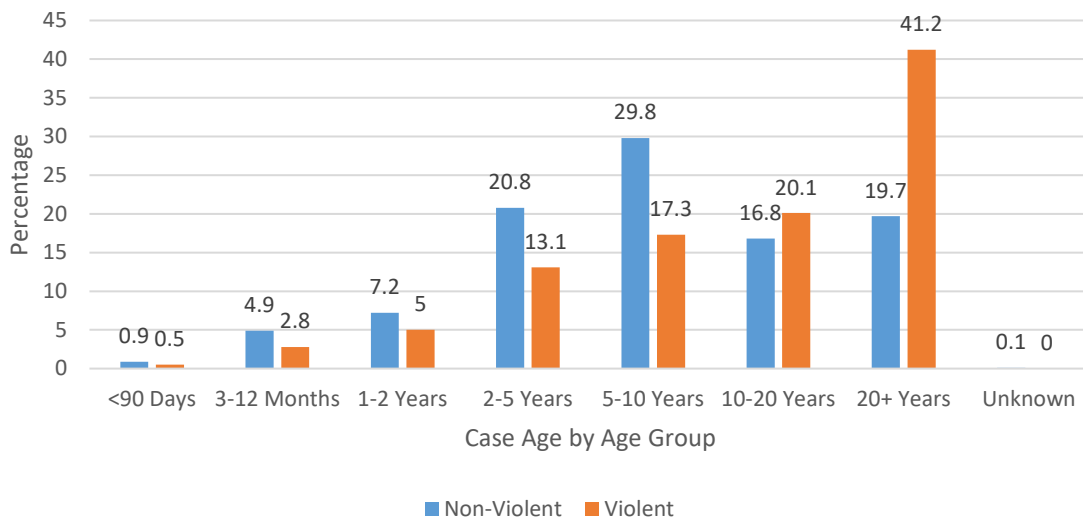
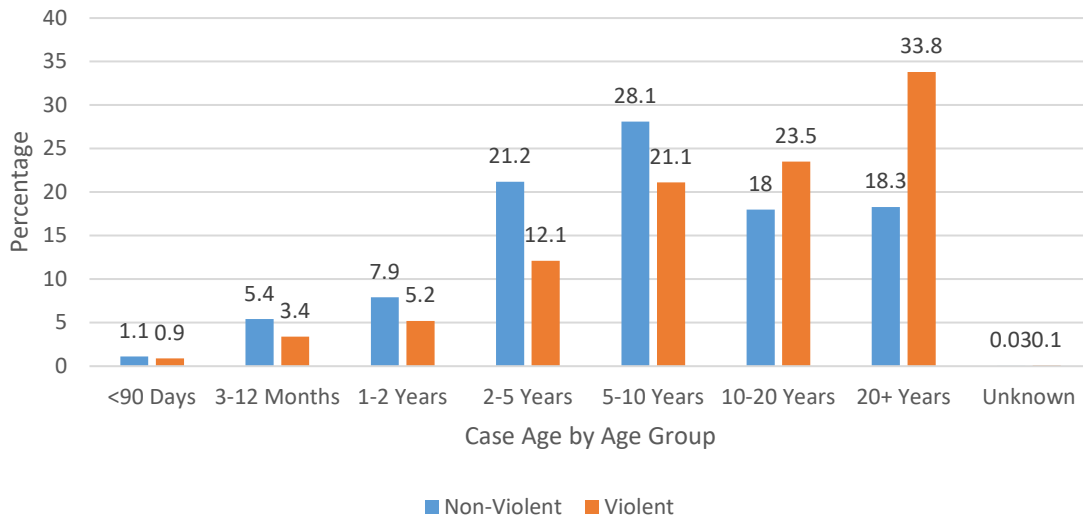


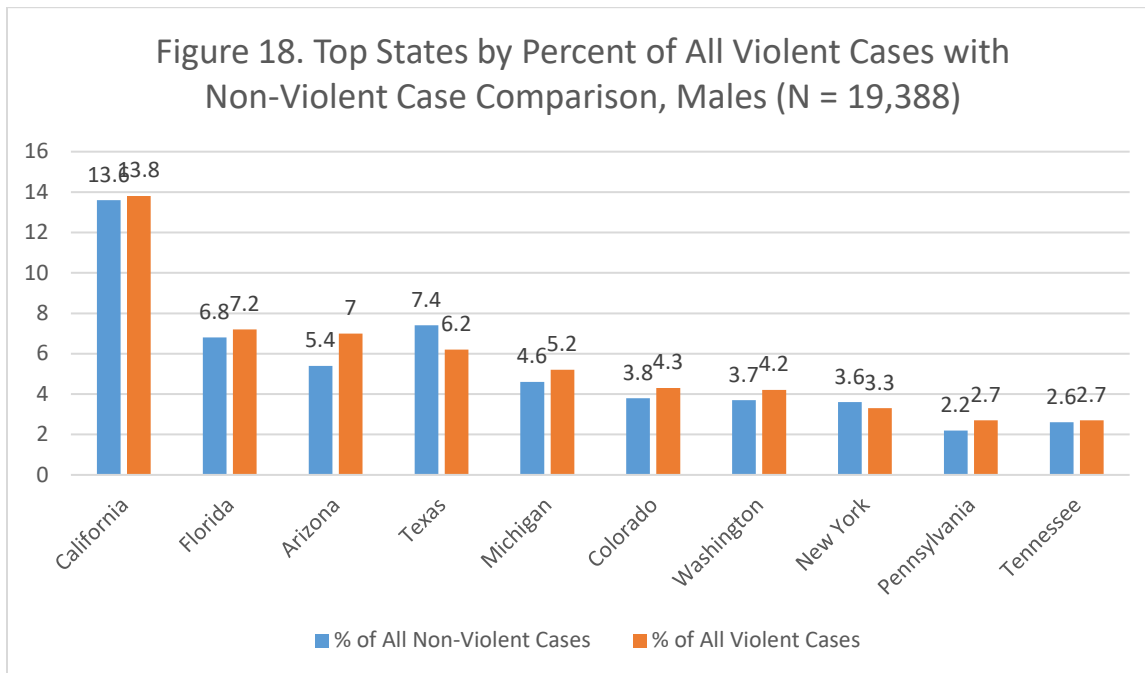
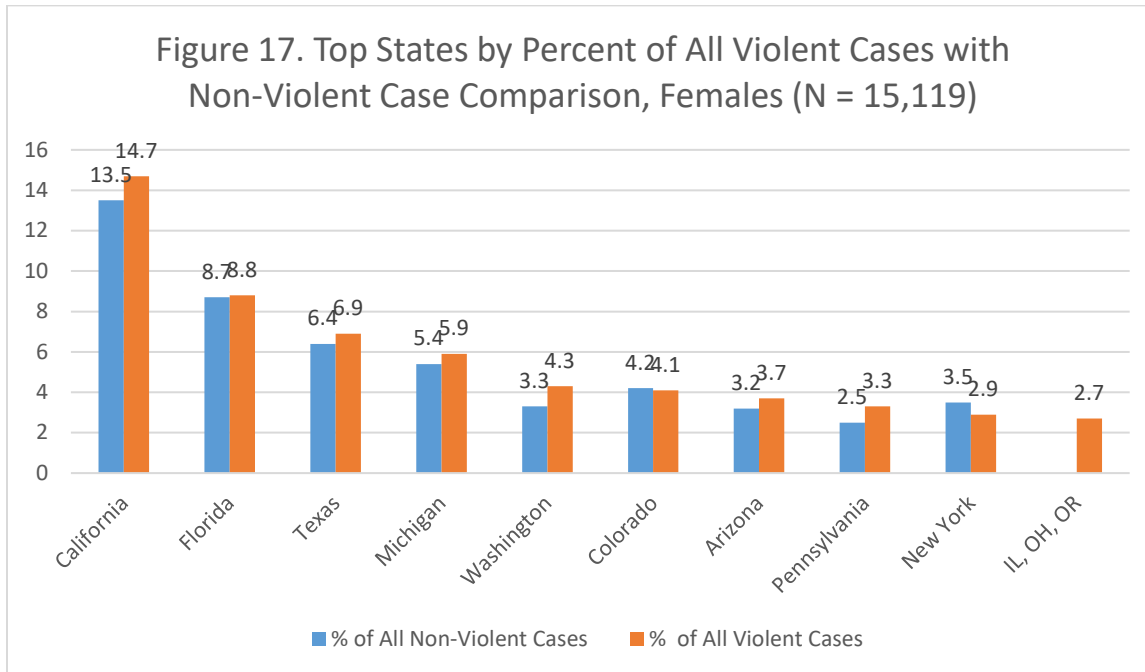
Figure 16. Comparison of Case Violence Across Age of Case Groups, Males (N = 19,388)



State Name (N = 34,507)

Figures 17 and 18 show the highest percentage of violent cases in absolute terms of case numbers (i.e., do not represent rates of violence), including the non-violent case percentage for comparison, by state. Among both females (Figure 17) and males (Figure 18) California had the highest percentage of all violent cases with 14.7% (n = 281) of all violent cases among females

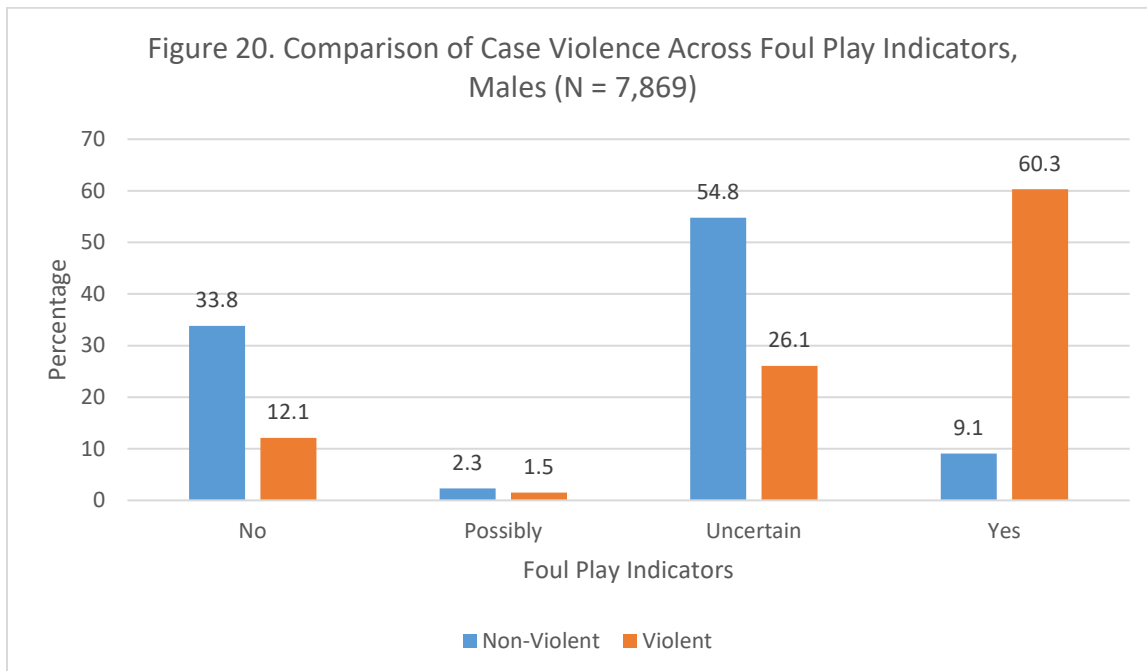
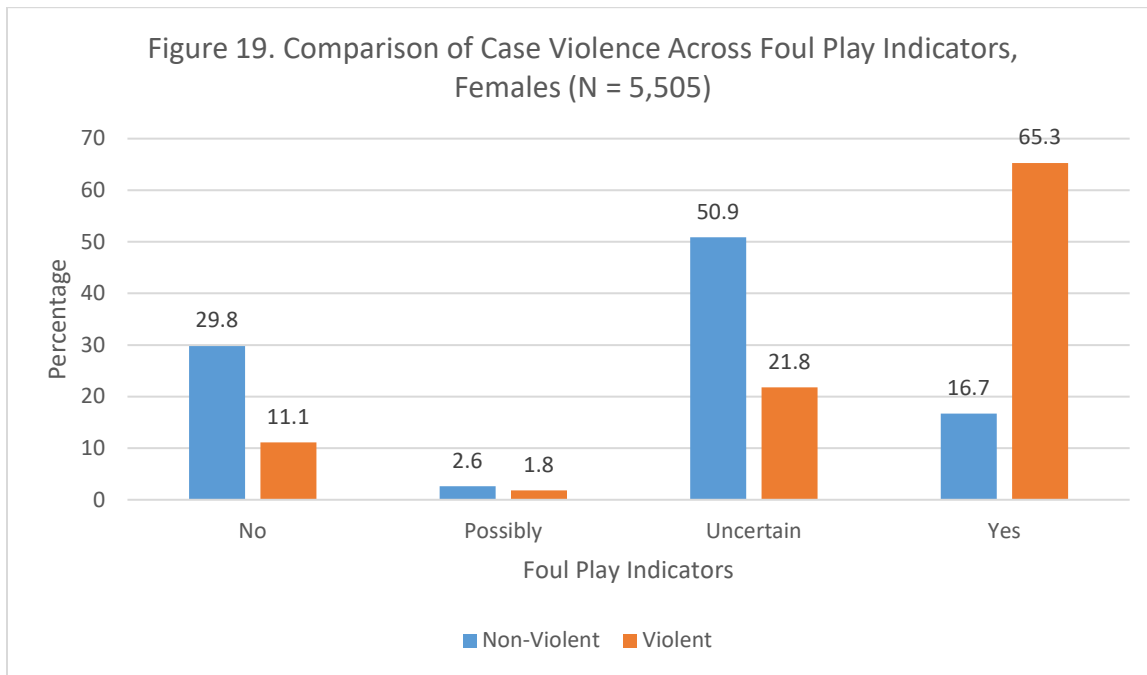
and 13.8% (n = 230) of all violent cases among males. These results may be due simply to the fact that the states represented enter more data than other states into the NamUs database.



Foul Play (N = 13,374)

Among female non-violent cases, 50.9% (n = 2,306) indicated uncertainty if foul play was likely (Figure 19). However, among female violent cases, 65.3% (n = 637) indicated likely foul play. A high proportion of cases not classified as violent did indeed have the foul play indicator checked (16.7%, n = 758). Among male non-violent cases, 54.8% (n = 3,850) indicated

uncertainty as to the presence of foul play (Figure 20). Among male violent cases, 60.3% (n = 509) indicated likely foul play. Again, a high proportion of cases that were not classified as being violent had the foul play indicator checked (9.1%, n = 640).



Population/Ethnicity (N = 34,503)*

Among females, White/Caucasian persons accounted for 55.0% (n = 7,263) of non-violent and 63.5% (n = 1,216) of violent cases (Figure 21). Black/African American females accounted for

20.8% of non-violent and 14.2% of violent cases, Hispanic / Latino females accounted for 11.0% of non-violent and 9.3% of violent cases, and American Indian/Alaska Native females accounted for 2.0% of non-violent and 1.7% of violent cases. Among males, White/Caucasian persons accounted for 61.4% (n = 10,885) of non-violent cases and 59.2% (n = 985) of violent cases (Figure 22). These results are to be expected as the majority of persons in the NamUs database are White/Caucasian. Black/African American males accounted for 15.5% of non-violent and 15.0% of violent cases, Hispanic / Latino males accounted for 9.5% of non-violent and 10.3% of violent cases, and American Indian/Alaska Native males accounted for 2.1% of non-violent and 2.2% of violent cases.

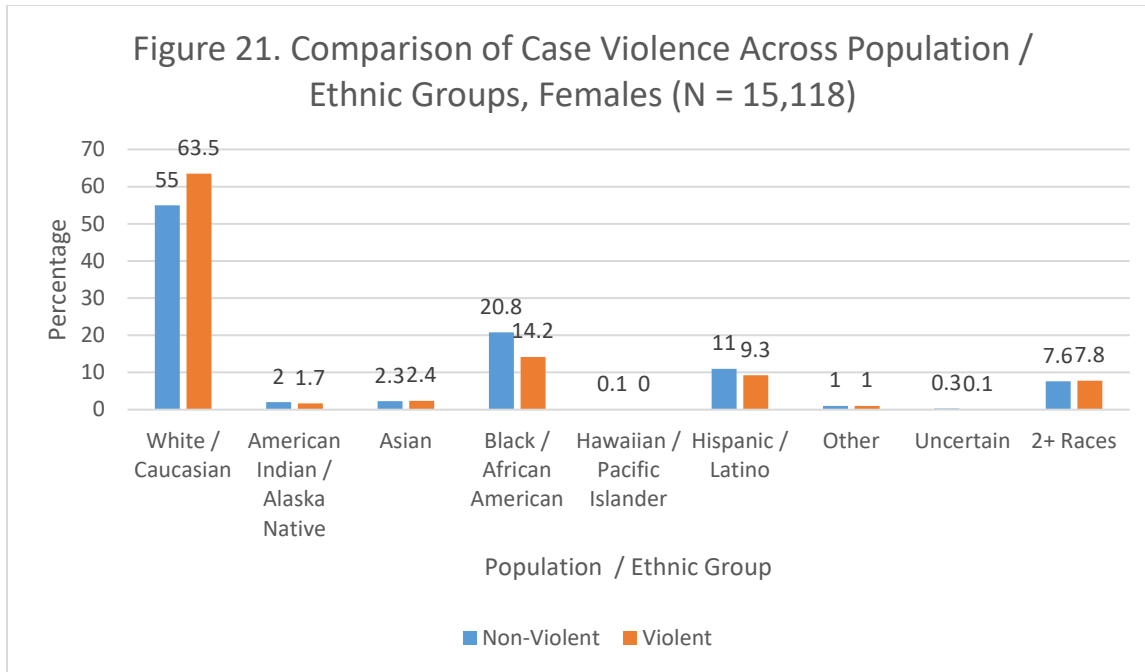
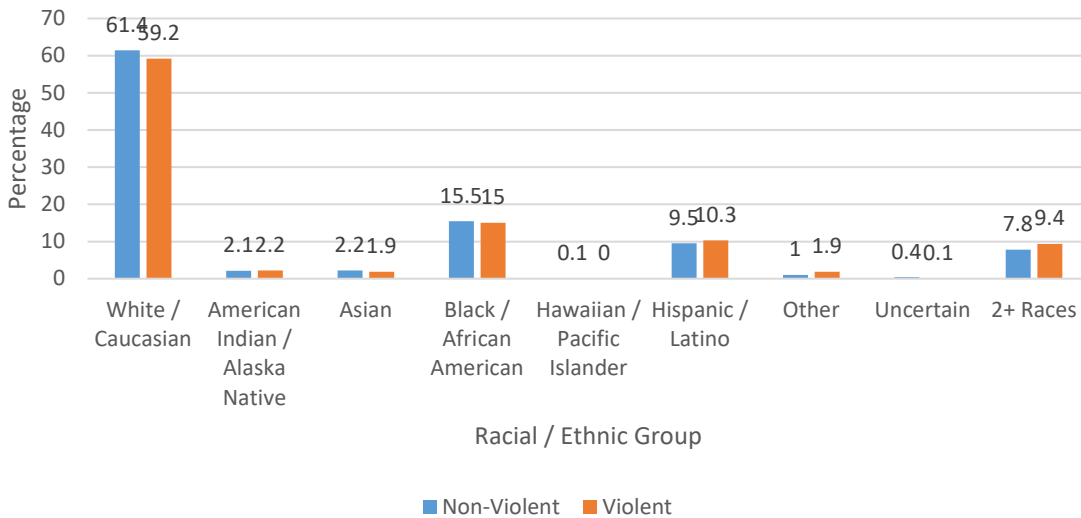


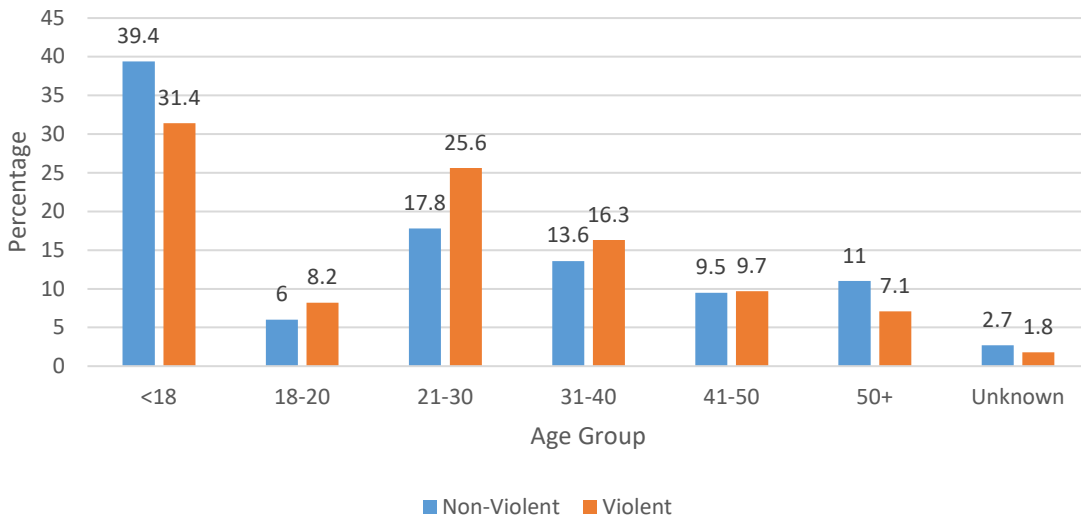
Figure 22. Comparison of Case Violence Across Racial / Ethnic Groups, Males (N = 19,385)

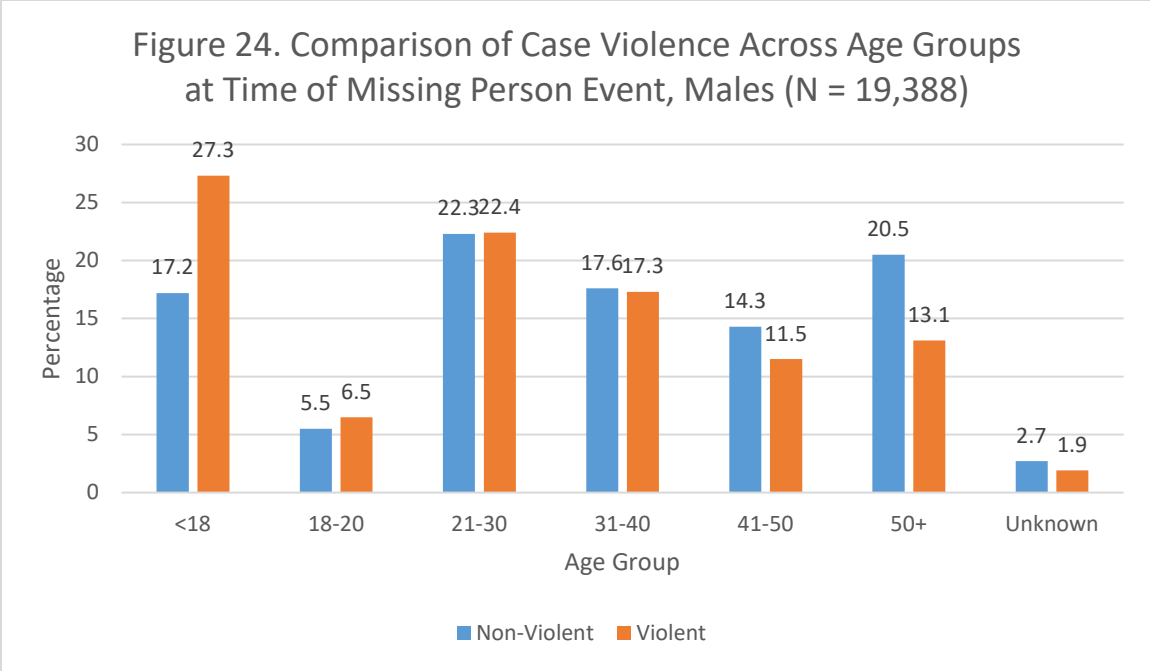


Age Missing (Group) (N = 34,507)

Among females, the highest percentage of cases were the under 18 age group for both non-violent (39.4%, n = 5,207) and violent (31.4%, n = 602) categories (Figure 23). Among males, the highest proportion of non-violent cases was the 21-30 age group (22.3%, n = 3,948), and the highest proportion of violent cases was the under 18 age group (27.3%, n = 454) (Figure 24)

Figure 23. Comparison of Case Violence Across Age Groups at Time of Missing Person Event, Females (N = 15,119)

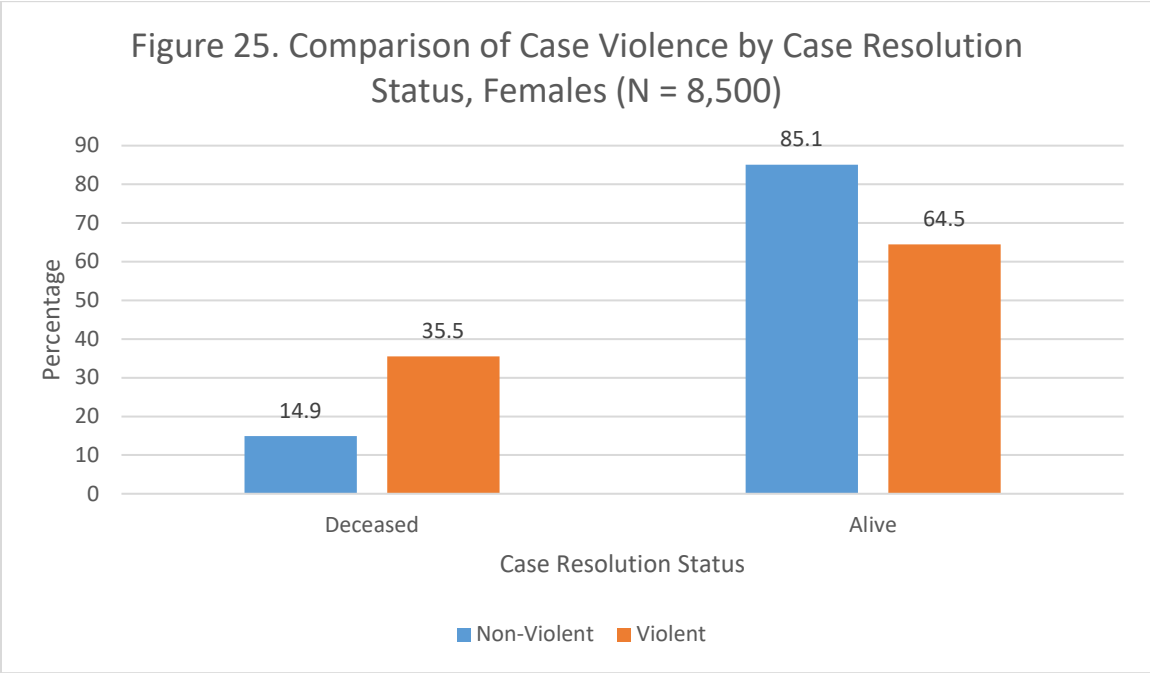


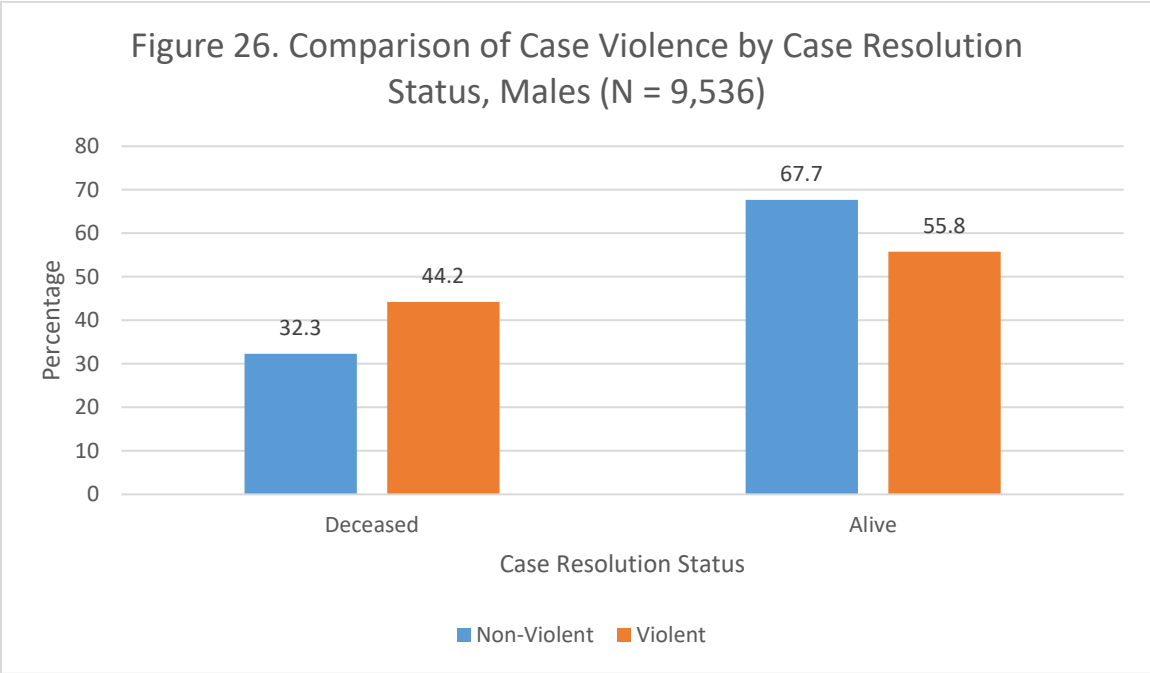


Resolved Cases (N = 18,051)

Case Resolution Status (N = 18,036)

Among females, 85.1% (n = 6,624) of non-violent cases and 64.5% (n = 465) of violent cases had a resolution where the individual was found alive (Figure 25). Among males, 67.7% (n = 6,031) of non-violent cases and 55.8% (n = 353) had a resolution where the missing person was found alive (Figure 26).





NamUs Assisted (N = 17,950)

Among females, 9.5% (n = 738) of non-violent cases and 15.8% (n = 113) of violent cases indicated NamUs assistance (Figure 27). For males, 15.8% (n = 1,402) of non-violent cases and 16.6% (n = 104) of violent cases indicated NamUs assistance (Figure 28).

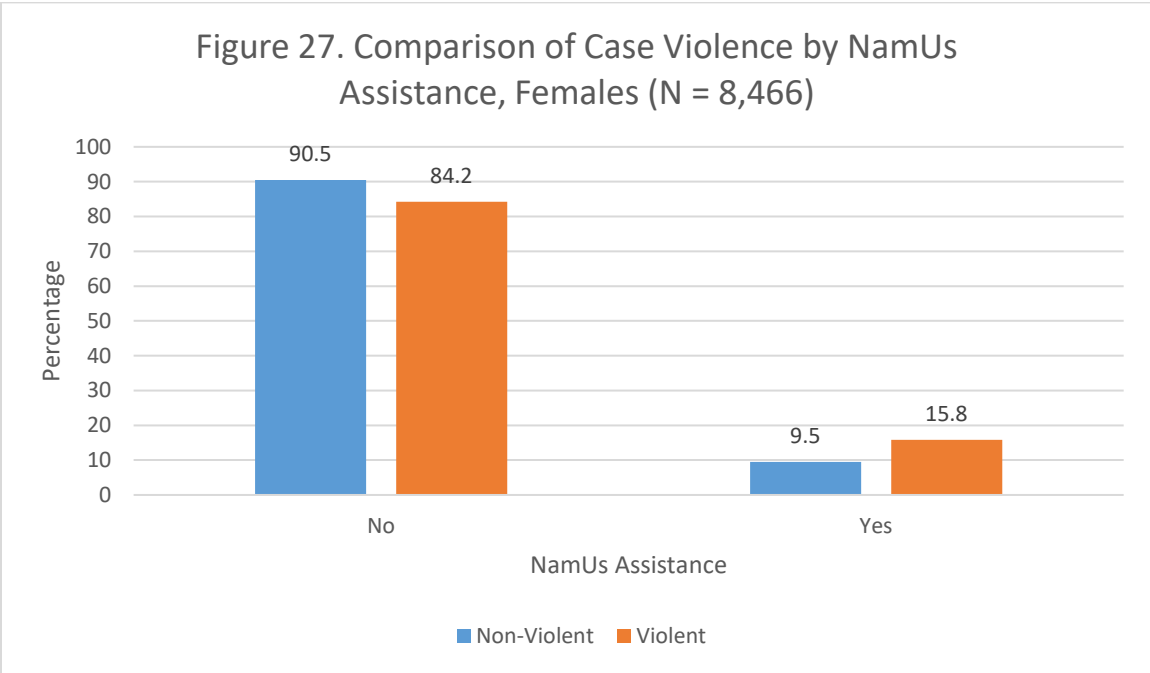
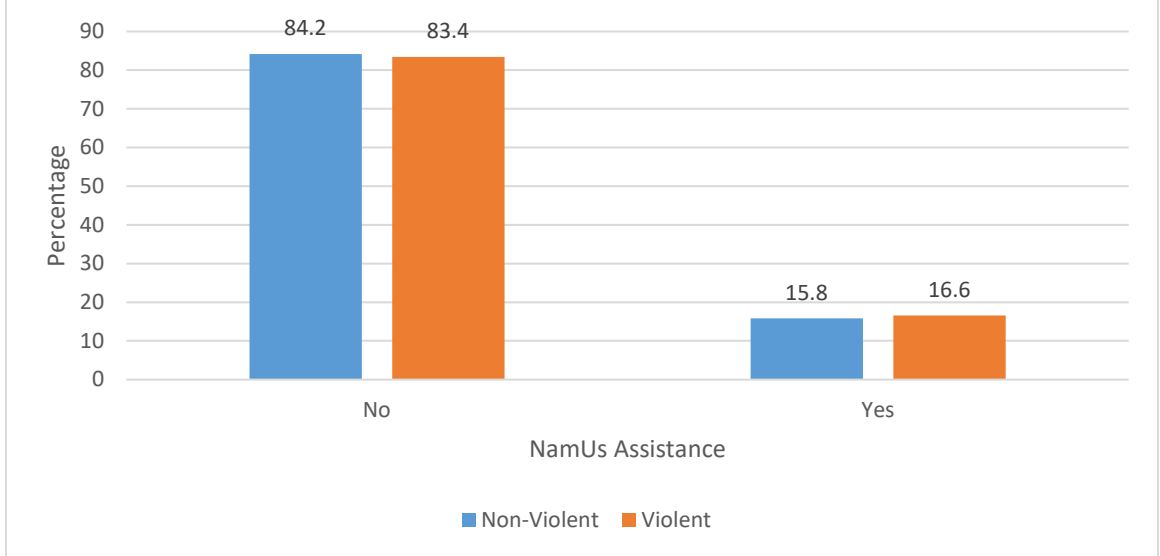


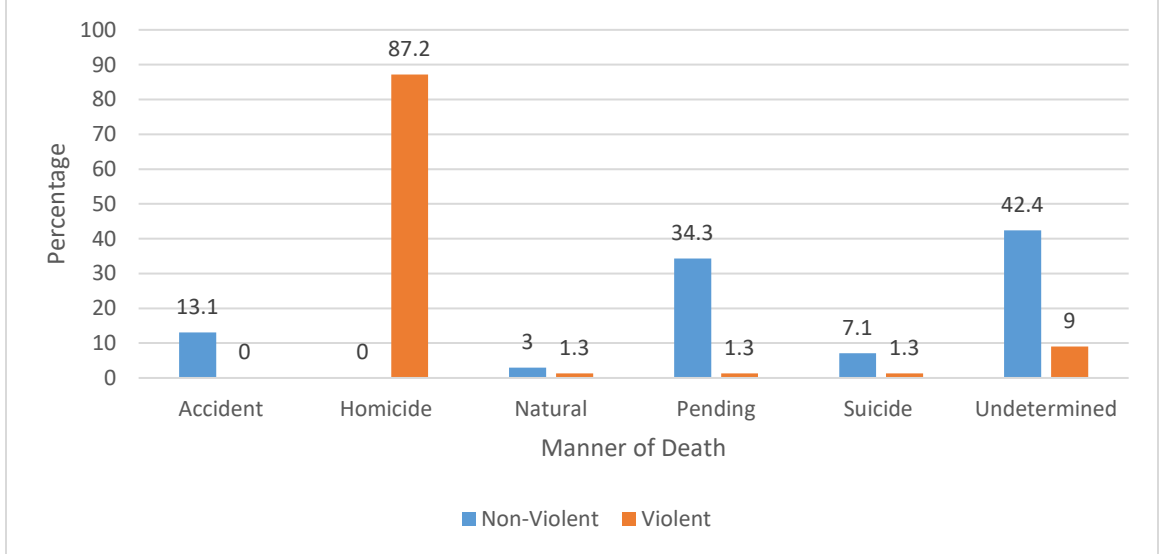
Figure 28. Comparison of Case Violence by NamUs Assistance, Males (N = 9,484)

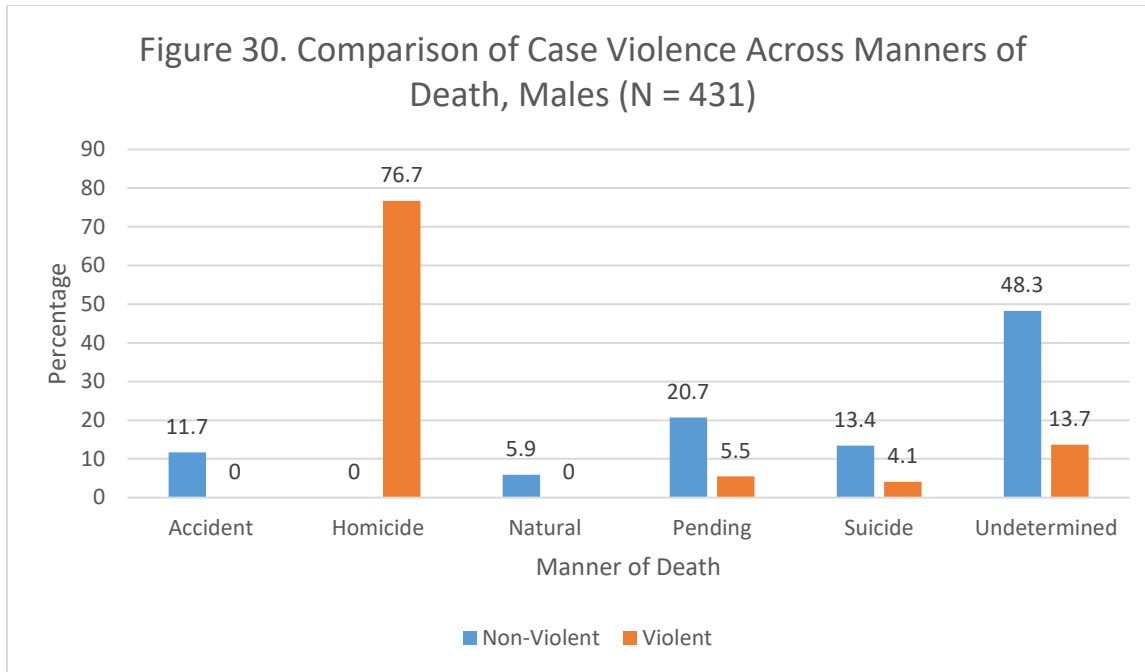


Manner of Death (N = 608)*

Among female cases classified as non-violent, the highest proportion of cases had undetermined as the MOD (42.4%, n = 42). Among female violent cases, 87.2% (n = 68) of cases were ruled as homicides, representing the highest proportion of MOD (Figure 29). Among male non-violent cases, the highest proportion for MOD was undetermined (48.3%, n = 173). For male violent cases, the highest proportion of MOD was homicide (76.7%, n = 56) (Figure 30).

Figure 29. Comparison of Case Violence Across Manners of Death, Females (N = 177)

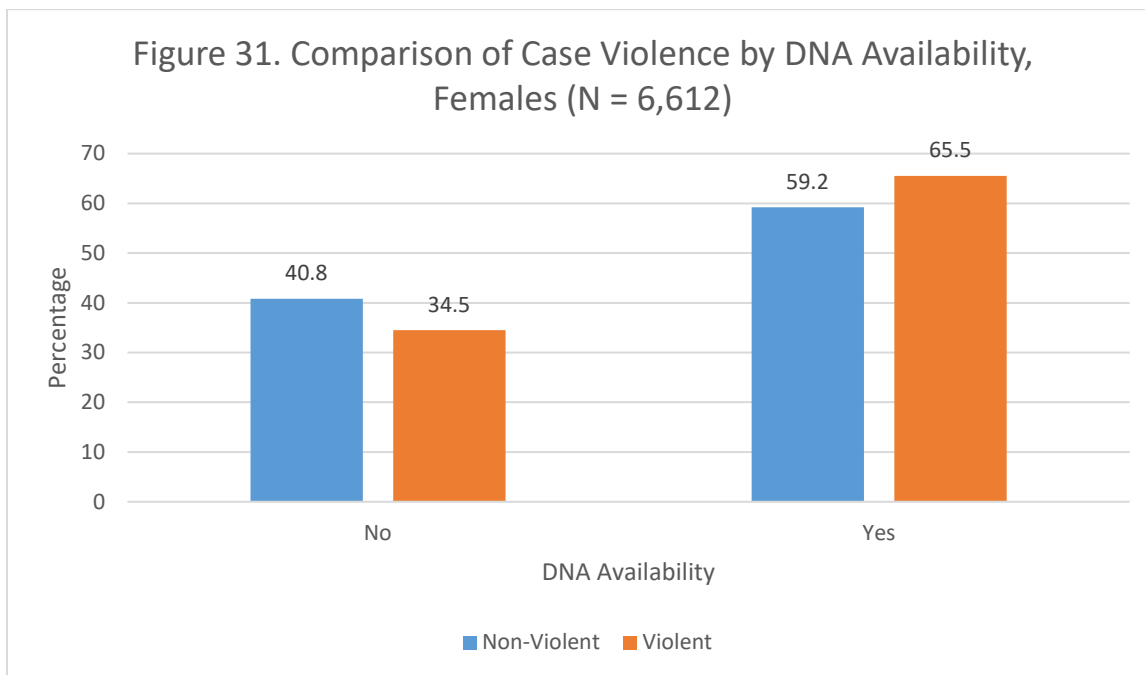


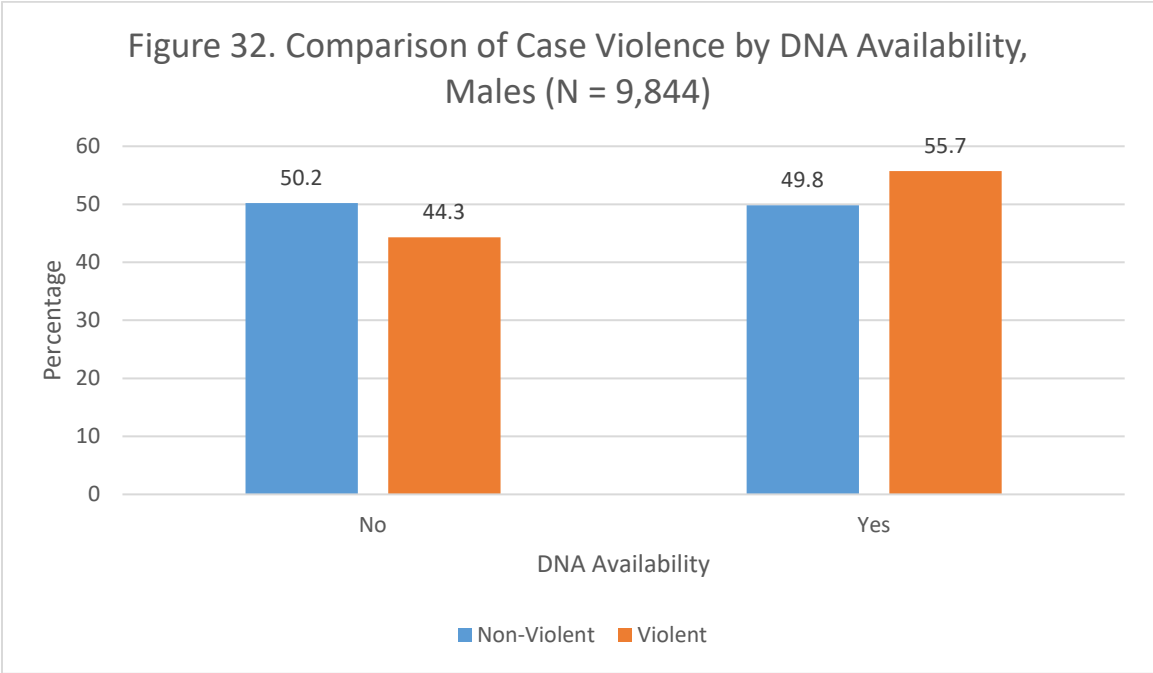


Unresolved Cases (N = 16,456)

DNA Availability (N = 16,456)

Among female non-violent cases, 59.2% (n = 3,207) of cases had DNA samples available, while 65.5% (n = 781) of violent cases also had such information available (Figure 31). Among males, 49.8% (n = 4,392) of non-violent cases and 55.7% (n = 574) of violent cases had DNA samples available for analysis (Figure 32).





Recommendations

The following recommendations are made based on the results from the case classification and analysis. The recommendations serve to improve the designation of violent cases, as well as in the investigation and resolution of cases involving violence based on the unique features that were uncovered. Additionally, these recommendations provide detailed suggestions for systems improvement and expansion for future research, evaluation, and analysis efforts. Overall, these recommendations could expand the NamUs mission by bringing together improved information gathering, technology enhancements, and human services to better resolve cases involving violence.

Figures 1 and 2 show that a higher proportion of female cases (12.7%) than male cases (8.6%) involved any type of violence.

- **Recommendation 1:** Implement discrete violence fields similar to that of the “foul play” field that can be checked by the case entrant to identify the case as possibly involving violence. Such fields may refer to specific types of violence identified in the text mining component of this project such as “fight,” “trafficking,” or other common keywords.
- **Recommendation 2:** The Victim Services Division should conduct a discovery phase to determine what victim services are appropriate for MP victims of violence or for family and friends of UP cases involving violence in order to more fully be able to address the needs of such persons.

Figures 15 and 16 show that the majority of violent cases for both females and males fell within the 20+ year age case group (41.2% female, 33.8% male)

- **Recommendation 3:** For long-term cold cases (20+ years), additional investigation or inquiry into the potential presence of violence should be undertaken by the assigned RPS or another member of the NamUs team.

Figures 19 and 20 show that there are cases which were classified by us as violent, but which were not indicated as having foul play (11.1% of female violent cases, 12.1% of male violent cases). Alternatively, there were cases in which the case was not classified as violent, but had the foul play indicator checked (16.7% of female non-violent cases, 9.1% of male non-violent cases). Further, Figures 19 and 20 show that a large proportion of non-violent cases do not have the indicator checked for foul play (16.7% females, 9.1% males). Additionally, overall, 21,133 cases (61% of all cases) did not provide any information for this field.

- **Recommendation 4:** For cases in which there is circumstantial text evidence of foul play, but the foul play indicator was not checked, or for cases in which there was no circumstantial text evidence of foul play, but the foul play indicator was checked, the assigned RPS should make efforts to obtain additional information from the case entrant prior to publishing the case.
- **Recommendation 5:** Given the high proportion of cases that did not have any foul play indicator checked, make this field mandatory to capture more comprehensive information on the potential burden of violence across all cases.
- **Recommendation 6:** If a case is marked as positive for foul play, then require the specific type of violence to also be checked referenced in Recommendation 1, and vice versa.

Figures 27 and 28 demonstrate a relatively low percentage of cases being assisted by NamUs for both males and females regardless of violent case status (non-violent: 9.5% female, 15.8% male; violent: 15.8% female, 16.6% male). More information is needed to determine if these data accurately reflect the value of NamUs or are an underrepresentation of the impact of NamUs.

- **Recommendation 7:** Require users at case close-out to indicate if NamUs assisted with case resolution. This prompt should capture the various ways in which NamUs may have helped in a case including through biometric services, analytical services, RPS assistance, among other areas in which NamUs serves to assist in case resolution.

Figures 29 and 30 demonstrate a large proportion of non-violent cases with either undetermined or pending MOD (76.7% combined females, 69.0% combined males).

- **Recommendation 8:** It is likely that some of these cases do involve violence or foul play. As such, a prompt or reminder message should be sent to the case entrant to ensure that the case file is fully filled out, including the foul play indicator and any associated circumstantial text.
- **Recommendation 9:** Have RPSs perform additional review of cases at close-out to have better information to analyze by ensuring all the fields are completed. In reviewing MOD information, it was noted that a large proportion of cases did not have final MOD information. Thus, NamUs staff should pay special attention to quality assurance before archival to have more accurate and complete data to develop best practices.

Conclusion

Overall, this study of MP case violence and associated characteristics provides a more detailed and nuanced insight into the impact and experiences of violence among MP within the NamUs database. The findings also help to detail concrete and distinct recommendations for systems expansion to better capture case violence and to better serve persons impacted by violent cases by identifying unique demographic and investigative characteristics of such cases that may aid in case investigation and resolution.. The key findings and results of the project also point to areas for further investigation where gaps in knowledge remain due to limitations of the NamUs data. However, in all, this report provides critical insights into case violence and the need for NamUs systems improvement that will lead to better services and system functionality for violent cases within the system and handled by staff and NamUs service components.

Limitations

The MP case classification and analysis had a number of limitations. First, NamUs is an operational database that is not intended to be a complete record of all investigative reports and notes. As such, some information necessary to carry out our analysis is lacking. Similarly, as an operational database, there is no standardized case entry as there might be with data collection through a typical survey data collection effort. Thus, there is variation in the type and amount of information provided for each case depending on several factors, including the integrity of any ongoing investigations, which may directly impact indicators of violence within NamUs.

Second, it is possible that there were deficiencies in our dictionary. All relevant keywords may not have been included, and some inflected forms may have been omitted. As such, violence and victimization may have been under-classified within WordStats.

Further, keywords were placed under broad violence types. It is possible that individual keywords were placed under inappropriate violence types. The classification of keywords into types was at the subjective interpretation of the primary coder in consultation with the secondary coder. However, it is still possible that alternative keyword classifications under different violence types might have affected the violence type results.

Additionally, secondary coding took place on only 10% of all cases, and tertiary coding occurred on only 20% of that 10% sample. These levels of second and tertiary coding may not have been high enough to truly assess and verify the dictionary in question. As such, deficiencies in the dictionary, whether the inclusion of inappropriate keywords or the exclusion of necessary words, may not have been adequately picked up using these coding sample thresholds.

Finally, final acceptance of results relied upon the inter-rater agreement between both the human coder and WordStat, as well as between two human coders. The level at which results were accepted may have been lower than was warranted to accept the results as true and valid. A higher threshold for inter-rater agreement may have impacted the results likely resulting in lower numbers of violent cases.

III. Unidentified Persons

Background

The issue of unidentified decedents presents a challenge ME/C and law enforcement (LE) agencies across the United States (Ritter, 2007). It is estimated that over 4,400 unidentified decedents are recovered each year with 1,000 of those decedents remaining unidentified after one year (Hickman, Hughes, Strom & Roper-Miller, 2007). In the same survey, the authors found that there were over 13,400 unidentified persons within ME/C offices. Hughes (2007) used the CDC's National Death Index (NDI) and found that between 1980 and 2004, there were 10,300 UPs within the index at some point during those years; however, it should be noted that the NDI is neither a national database for UP nor do states uniformly report on death certificates if a person's identity is unknown. As of October 28, 2020, there were 13,485 active UP cases within the NamUs database. Most of these NamUs UP cases are not known to be the result of foul play and thus tend to represent natural, accidental, suicidal, or undetermined MOD. Overall, the burden of UP cases on the ME/C system and to LE agencies is significant due, in part, to the volume of cases, as well as the lack of resources, both human and forensic, needed to attain positive identification (Ritter, 2007).

Violence and Unidentified Persons

Although the majority of NamUs UP cases involve non-violent MOD, a portion of UP cases do entail homicide as the MOD. As of October 28, 2020, 2,077 (15.4%) of active NamUs UP cases had been ruled homicides by ME/Cs. However, that number may be higher given that 5,724 cases were listed as having an undetermined MOD, which may include cases of homicide. In examining UPs within the National Crime Information Center, Hughes (2007) found that 27% of all UP cases had been ruled homicides. Hughes also found that the majority of deaths had an undetermined cause of death (49%), which is slightly higher than in NamUs (42.4%) as of October 28, 2020. It is likely that some of these undetermined cases were homicides, but not ruled as such due to the condition of the remains, the amount of remains recovered, or unknown circumstantial information. Regardless, the number of UP homicides is likely underestimated within NamUs.

Understanding UP Violence in NamUs

The current project sought to better understand the burden of violence in UP cases within a national database. To do so, it relied on case file information from active and archived cases within the database. Violence was defined in two ways. First, any case ruled as a homicide by a ME/C was coded as violent. Second, text mining of undetermined and pending manners of death, as well as cases for which no MOD was provided, was conducted to determine if physical aggression resulted in deadly harm to an individual. In completing this work, a broader, more nuanced picture as to the extent of violence associated with cases within the database was elicited, as well as a better understanding as to how violence may be related to other characteristics of a case.

Methods

Similar to the MP section, this section describes the methods used to complete case classification and analysis. A more detailed discussion of the methods, including more nuanced information on dictionary building, misspellings, coding, and reporting are provided in Appendix A – Technical Note.

Project Data

Data for this project came from the unidentified section of the NamUs database. Data include all published (active) and archived (resolved) cases as of the pull date of August 29, 2019. Data were pulled by the Senior Business Intelligence Developer and delivered to the project team using the secure, password-protected sharing service OneDrive. Data were maintained on an encrypted, password protected, UNTHSC-provided computer.

Data were delivered as a single Microsoft Excel file with three separate sheets and downloaded via OneDrive. Sheet one included Case Information that provided demographic and case details on the UP; sheet two included Circumstances that provided in text entries spaces for general circumstance information available for public view, Agency Notes available only for professional users, and Circumstances of Resolution that provided information as to the outcome of a resolved case, and cause of death information; and sheet three included Agency Information with detailed information on the case owner and any contributors. These three sheets were saved as separate Microsoft Excel files.

Case Classification

Initial Classification

Initially, the MOD provided by the ME/C was used to classify cases as violent or non-violent. All cases indicated as homicides were classified as violent while all other cases excluding cases where MOD determination was “pending” were classified as non-violent. For the study purposes, suicide was considered a non-violent MOD, as were accidental and natural MOD.

Initially, text mining classification of cases as violent or non-violent was similar to that of the MP cases and was used for cases indicated as having an undetermined MOD or for which no MOD was provided. In the final run of classification of the data, cases with a pending MOD were also mined using similar procedures as the procedures used for undetermined and not provided MOD cases. As with MP cases, WordStat 8.0.20 (Provalis Research; Montreal, Canada) was used for the classification of cases as violent or non-violent.

Classification Dictionary

Similar to the MP dictionary, to classify cases as violent or non-violent, a classification dictionary was created in WordStat that included keywords relevant to UP cases. This dictionary was initially based on the MP dictionary but was further developed and modified collaboratively by the project team to ensure the best coverage and most exhaustive list of potential keywords that were relevant to the classification of UP cases as violent or non-violent. In total, 37 words were identified and included in the initial dictionary. Keywords were not categorized into specific types of violence as with the MP dictionary; rather, each keyword was a standalone indicator of case violence. Amendments to the initial dictionary were based on the iterative classification outcomes. The final classification dictionary included 40 keywords and can be found in Appendix B.

Classification Runs

In total, there were three classification runs for UP cases. The order of steps and overall approach to the runs varied only slightly and was based on lessons learned from previous runs.

General Classification Approach. Runs 1 and 2 involved cases with an “undetermined” MOD or with a MOD that was not provided. Run 3 included only cases in which the MOD was “pending.” The same general process was followed for each classification run. First, data were delivered and, unless cleaned data were already available, data were cleaned, resulting in the removal of cases that did not have any information in the circumstance fields, that only indicated some form of “unknown,” or that contained only numbers or symbols. Cleaned data were then imported into WordStat for classification.

A misspellings check was also conducted that was completed using WordStat’s built-in functionality. Each potential misspelling was checked using a “keyword-in-context” feature to ensure the misspelling truly represented a misspelled word, a proper noun, or some other otherwise correct spelling for a different word. The final run of the data included corrected misspelled words. Overall, this final classification resulted in cases being coded as violent or non-violent by WordStat for analysis.

Following each run, data were imported into Stata, and a random sample of 10% was drawn for second coding. Percent agreement was calculated between WordStat and the human coder. Following the calculation of percent agreement, a disagreements check was performed to understand discrepancies in coding between WordStat and the human coder. For Run 2, percent agreement was calculated to be 88.2%. For Run 3 with just “pending” MOD case, percent agreement was 82%. All “pending” disagreements were instances in which the computer found a keyword and coded a case as violent, but in which the human coder did not find the case to indicate potential or actual violence. As such, these cases were removed during the final violence coding procedures described below.

When appropriate, tertiary coding was conducted by a second human coder to check the coding of the first human coder. To do so, a subsample of 20% of the 10% sample drawn for second coding was drawn at random in Stata. Tertiary coding also resulted in a percent agreement between human coders resulting in 97.8% agreement for Run 2 and 95.2% agreement for Run 3 with “pending” cases only.

Final Violence Coding. Following secondary and tertiary coding, and upon findings of adequate percent agreement for both secondary and tertiary coding, a manual coding of all cases indicated as violent by WordStat was conducted to verify the violent nature of the case. These final codings represent final case violence for the purposes of further analysis. During this process, in Run 2, 96 cases were found to be valid cases of violence. In Run 3, an additional 14 cases were found to be valid cases of violence. These 110 cases were then manually re-coded within the data to be violent.

Case Analysis

Following completion of case classification, the data were prepared for analysis such that meaningful comparisons could be made between violent and non-violent cases for both female and male cases.

Final Dataset Creation

To arrive at a final analytic dataset, all cases in which violence was present based on classification procedures were changed to indicate such violence by switching the mOD

classification to “homicide.” In all, there were 16,857 cases available for analysis, of which 2,655 indicated violence. For analysis, which was stratified by gender, cases with no gender provided (n = 7) or unsure gender (n = 1,009) were dropped resulting in a final case count of 15,841.

Variables

Variables used in the analysis stage of this project included violence variables, case characteristics, demographic characteristics, and biometric characteristics. Each set of variables is described below.

All Case Variables. All case variables included demographic characteristics and general case characteristics as displayed in Table 1.

Table 1. All Case Variables (N = 15,841)

Variable	Definition	Code	N
Final Violence	Indicate whether or not a case did or is suspected of involving foul play or violence	0 = No	13,186
		1 = Yes	2,655
			Total: 15,841
SexName2	Indicates initial biological gender determination as defined by reporting party	0 = Female	3,458
		1 = Male	12,383
			Total: 15,841
RaceEthnicity2	Indicates initial race/ethnicity determination as defined by reporting party	0 = White / Caucasian	6,239
		1 = American Indian / Alaska Native	102
		2 = Asian	265
		3 = Black / African American	2,555
		4 = Hispanic / Latino	1,877
		5 = Other	178
		6 = Uncertain	2,666
		7 = 2+ Races	1,928
			Total: 15,810
Age Group	Estimated age group of decedent based on median age of decedent	0 = <18	645
		1 = 18-20	330
		2 = 21+	12,677
			Total: 13,652
StateName2	State where person in case was found deceased	1 = Alabama	78
		2 = Alaska	54
		3 = Arizona	2,161
		4 = Arkansas	130
		5 = California	2,833
		6 = Colorado	115
		7 = Connecticut	85
		8 = Delaware	42
		9 = DC	51
		10 = Florida	1,162
		11 = Georgia	285

12 = Guam	2
13 = Hawaii	64
14 = Idaho	23
15 = Illinois	249
16 = Indiana	78
17 = Iowa	19
18 = Kansas	30
19 = Kentucky	113
20 = Louisiana	172
21 = Maine	21
22 = Maryland	367
23 = Massachusetts	240
24 = Michigan	475
25 = Minnesota	58
26 = Mississippi	60
27 = Missouri	144
28 = Montana	21
29 = Nebraska	8
30 = Nevada	325
31 = New Hampshire	8
32 = New Jersey	436
33 = New Mexico	161
34 = New York	1,589
35 = North Carolina	157
36 = North Dakota	3
37 = Ohio	171
38 = Oklahoma	243
39 = Oregon	218
40 = Pennsylvania	363
41 = Puerto Rico	235
42 = Rhode Island	21
43 = South Carolina	74
44 = South Dakota	7
45 = Tennessee	179
46 = Texas	1,926
47 = Utah	50
48 = Vermont	22
49 = Virgin Islands	3
50 = Virginia	198
51 = Washington	169
52 = West Virginia	46
53 = Wisconsin	75
54 = Wyoming	17
	Total: 15,840

ConditionOfRemainsName2	Indicates extent to which remains had a recognizable face	0 = Not recognizable	11,633
		1 = Recognizable face	3,891
			Total: 15,524
Age of Case Groups2	Indicates period of time between when decedent's body was found and present day	0 = 0-1 Years	234
		2 = 1-5 years	2,265
		3 = 5-10 years	2,404
		4 = 10-20 years	3,641

5 = 20+ years 7,296
6 = Unknown 1
Total: 15,841

Resolved Case Variables. Resolved case variables include characteristics related to the case investigation or to outcomes of the case as shown in Table 2.

Table 2. Resolved Case Variables (N = 3,605)

Variable	Definition	Code	N
NamUsAssisted	Indicator defined by professional user at case close-out to denote if NamUs was useful or an important factor in case resolution	0 = No 1 = Yes	615 1,556 Total: 2,171
MannerOfDeathName3	Indicates manner of death as determined by a medical examiner or coroner	1 = Accident 2 = Homicide 3 = Natural 4 = Pending 5 = Suicide 6 = Undetermined	438 566 239 469 114 1,043 Total: 2,869
MethodOfId_Anthropology	For resolved cases, indicates if anthropology was a method used to make positive identification	0 = No 1 = Yes	3,320 273 Total: 3,593
MethodOfId_CircumstantialInforma	For resolved cases, indicates if circumstantial information was a method used to make positive identification	0 = No 1 = Yes	3,184 409 Total: 3,593
MethodOfId_Dental	For resolved cases, indicates if dental comparison was a method used to make positive identification	0 = No 1 = Yes	3,268 325 Total: 3,593
MethodOfId_Fingerprints	For resolved cases, indicates if fingerprint comparison was a method used to make positive identification	0 = No 1 = Yes	2,650 943 Total: 3,593
MethodOfId_MtDNA	For resolved cases, indicates if mitochondrial DNA was a method used to make positive identification	0 = No 1 = Yes	2,444 1,149 Total: 3,593
MethodOfId_NucDNA	For resolved cases, indicates if nuclear DNA was a method used to make positive identification	0 = No 1 = Yes	2,049 1,544 Total: 3,593
MethodOfId_Other	For resolved cases, indicates if other means were a method used to make positive identification	0 = No 1 = Yes	3,572 21 Total: 3,593
MethodOfId_Radiograph	For resolved cases, indicates if radiographs were a method used to make positive identification	0 = No 1 = Yes	3,522 71 Total: 3,593

MethodOfId_VisualIdentification	For resolved cases, indicates if visual identification was a method used to make positive identification	0 = No 1 = Yes	3,223 370 Total: 3,593
---------------------------------	--	-------------------	------------------------------

Unresolved Case Variables. Unresolved case variables included a single biometric characteristic indicating if DNA had been uploaded and the analysis completed as shown in Table 3.

Table 3. Unresolved Case Variables (N = 12,236)

Variable	Definition	Code	N
DNA_Indicator2	Indicates if at least one DNA sample analysis has been completed	0 = No 1 = Yes	5,965 6,271 Total: 12,236

Analytic Approach

Similar to the MP analysis, analysis of the final cleaned case data proceeded in various steps. First, univariate descriptive statistics for each variable were completed as reported in the variable tables. Then, univariate statistics stratified by missing person gender were run for all violence variables. Finally, gender stratified (male, female) bivariate statistics comparing final case violence (yes, no) across levels of each remaining variable were run in order to assess differences in each variable by the presence or absence of case violence for each of the genders.

For each variable, all available cases with information on the variable were included for analysis. As such, many variables represent different sample sizes and samples. Further, analyses were separated into variables appropriate for all cases, variables for resolved cases only, and variables for unresolved cases. Resolved and unresolved status was determined by the lifecycle status of the case as either archived (resolved) or published (unresolved) as of the date of the data pull.

Results

Overall, there were 15,841 valid cases for analysis. Demographically, 3,458 (21.8%) cases were female, and 12,383 (78.2%) cases were male. Figures 33-52 show results for the comparisons made in the analysis. An asterisk (*) indicates that at least one cell count was less than 10 cases.

All Cases (N = 15,841)

Final Violence (N = 15, 841)

Among females, 27.2%, (n = 941) experienced a violent death (Figure 33). The proportion of males who experienced a violent death was lower than that for females (13.8%, n = 1,714) (Figure 34).

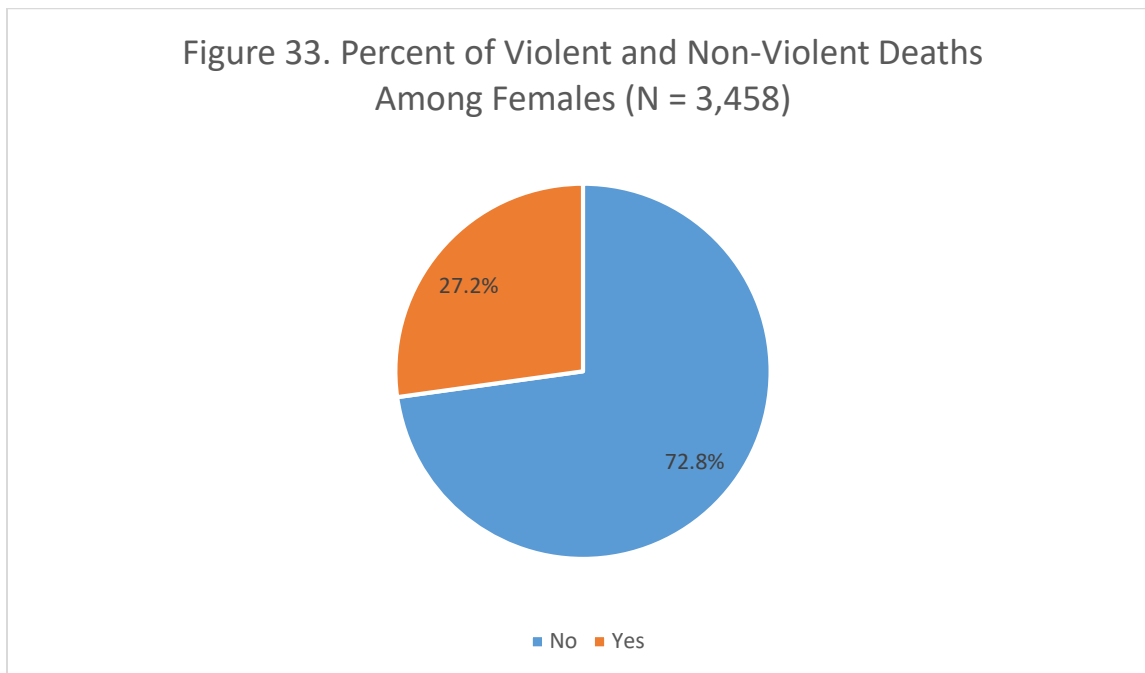
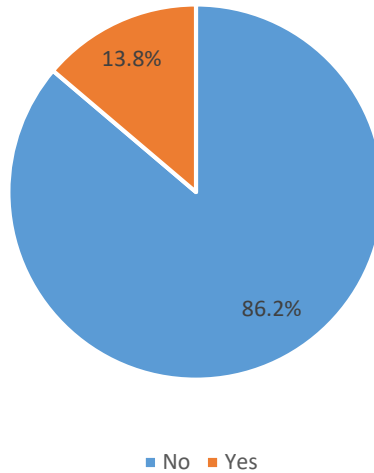


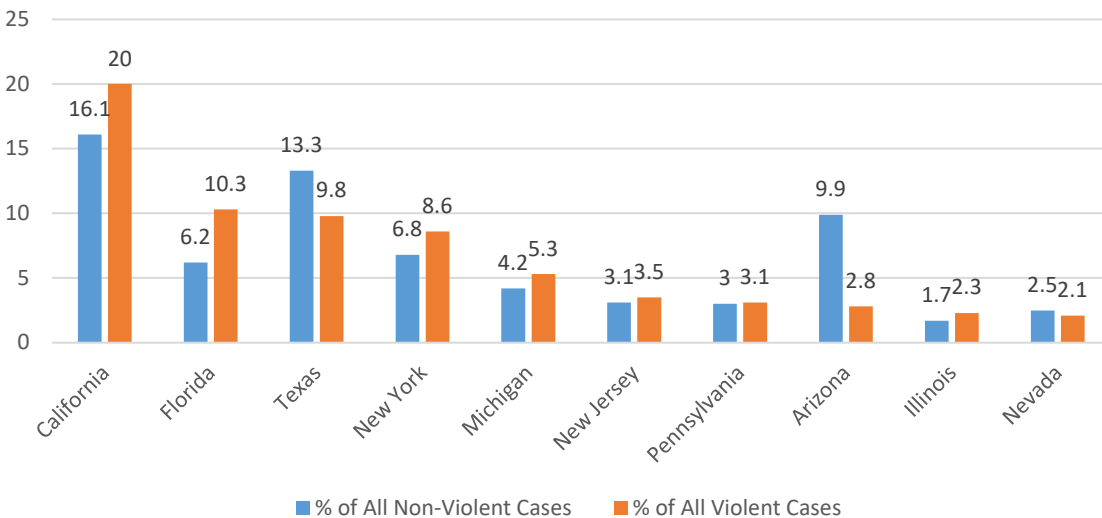
Figure 34. Percent of Violent and Non-Violent Deaths Among Males (N = 12,383)

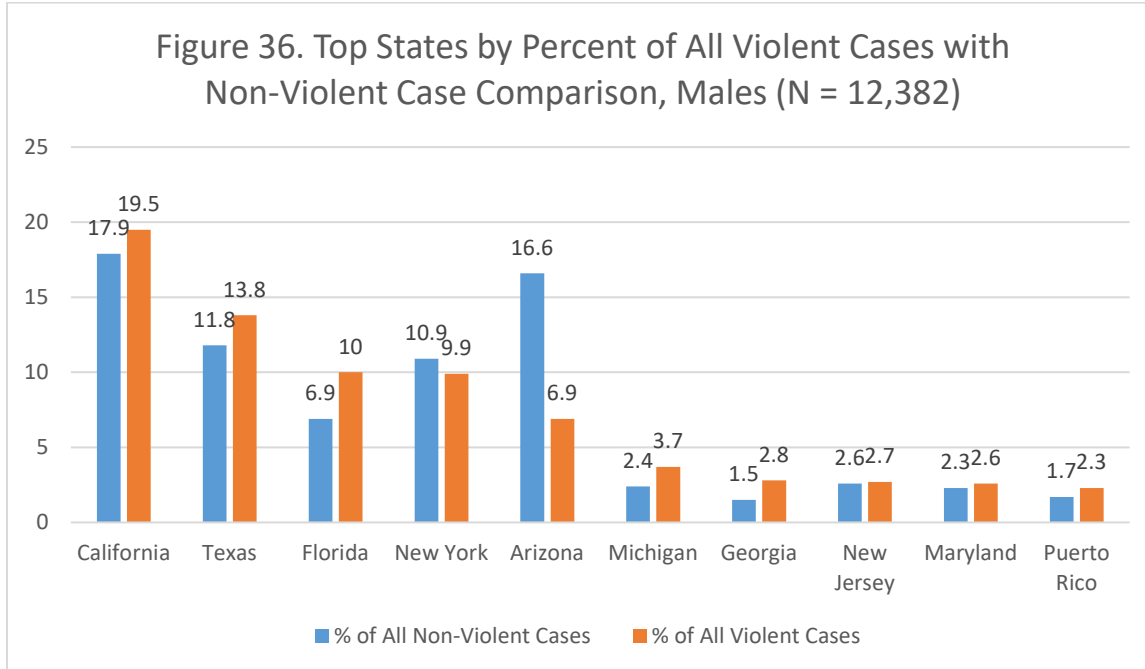


State (N = 15,840)

Figures 35 and 36 display the states with the highest percentage of violent cases with their non-violent case percentages presented for comparison. These numbers represent absolute percentages, and do not represent rates. As such, the percentages displayed may represent participation in NamUs as opposed to actual higher percentages per capita among the states shown. Among both females (Figure 35) and males (Figure 36), California is the top state displaying the most violent cases with the state representing 20.0% (n = 188) of all violent cases among females and 19.5% (n = 335) of all violent cases among males.

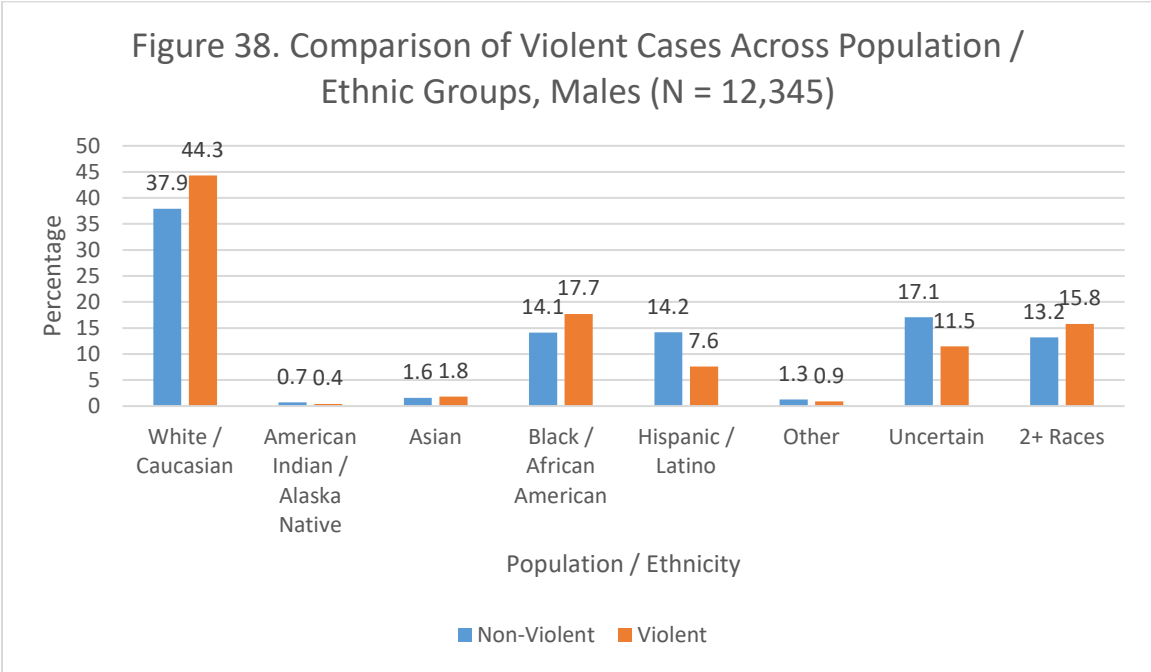
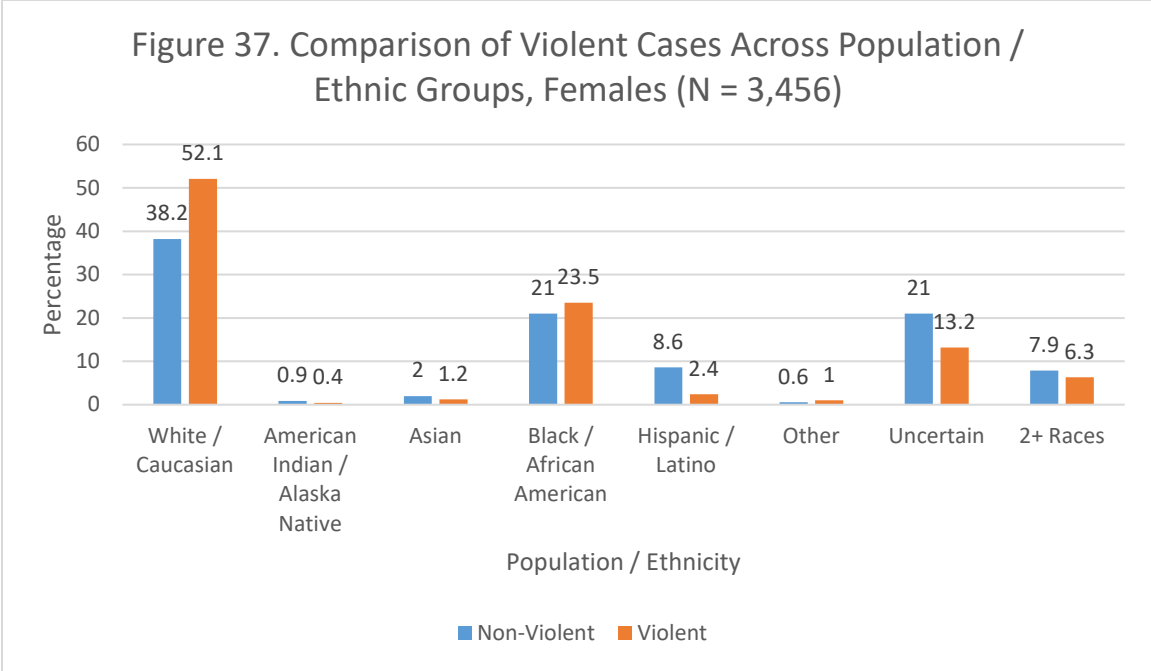
Figure 35. Top States by Percent of All Violent Cases with Non-Violent Case Comparison, Females (N = 3,458)





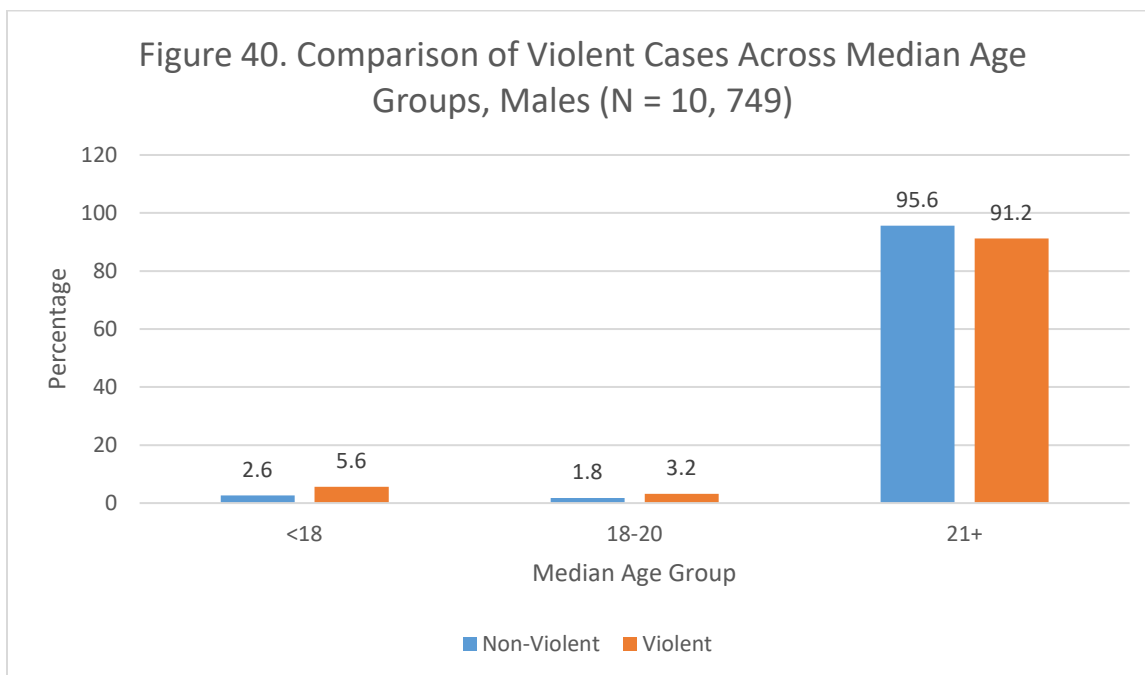
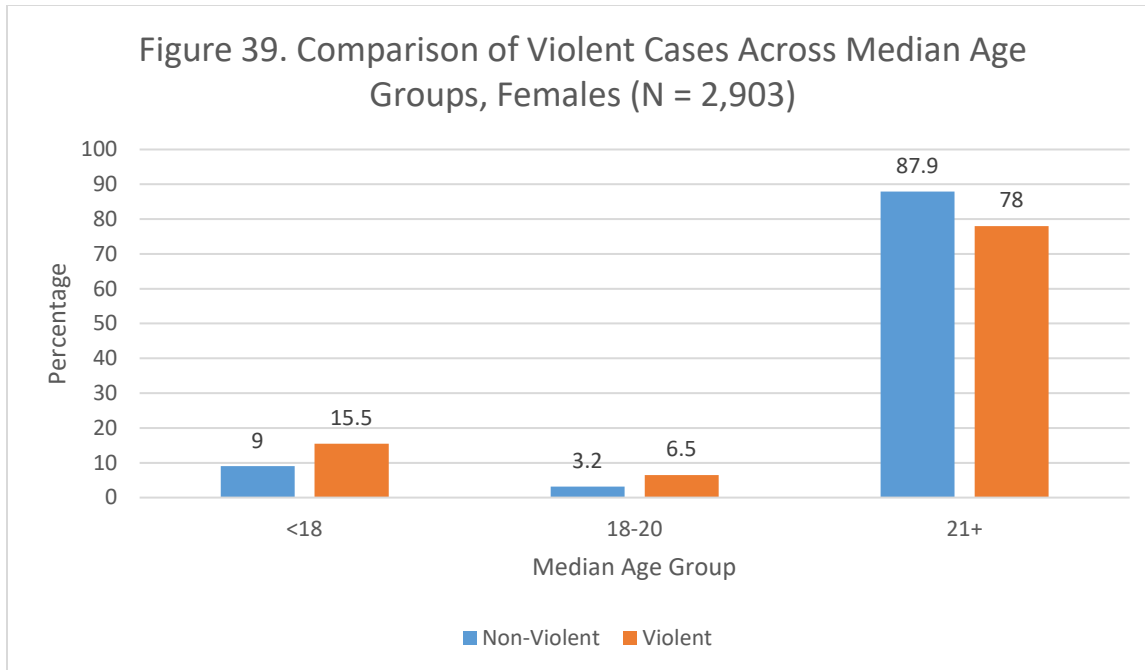
Population/Ethnicity (N = 15,810)*

Among female non-violent cases, 38.2% (n = 960) were White/Caucasian individuals, while 52.1% (n = 490) of female violent cases were among the same demographic (Figure 37). Black / African Americans made up 21.0% of non-violent and 23.5% of violent cases, Hispanic / Latino females made up 8.6% of non-violent and 2.4% of violent cases, and American Indian / Alaska Native females made up 0.9% of non-violent cases and 0.4% of violent cases. Among male non-violent cases, 37.9% (n = 4,030) of cases were among White/Caucasian individuals while among male violent cases, 44.3% (n = 759) were among White/Caucasian individuals (Figure 38). Black / African Americans made up 14.1% of non-violent and 17.7% of violent cases, Hispanic / Latino males made up 14.2% of non-violent and 7.6% of violent cases, and American Indian / Alaska Native males made up 0.7% of non-violent cases and 0.4% of violent cases. These results are not surprising given that White/Caucasian individuals make up the majority of all cases.



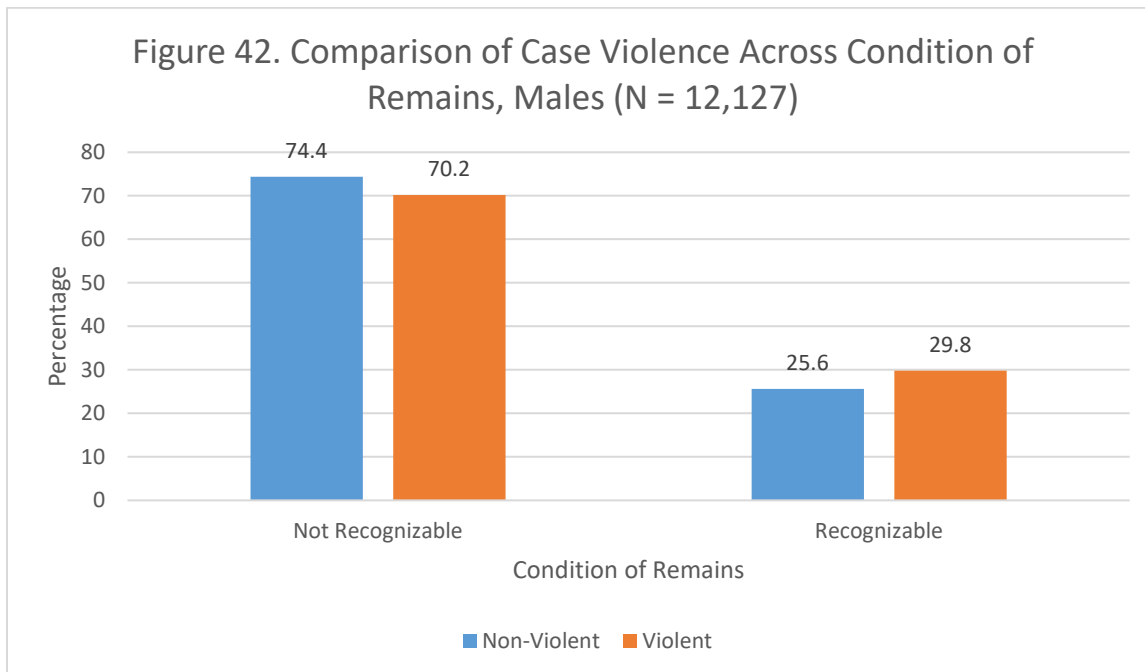
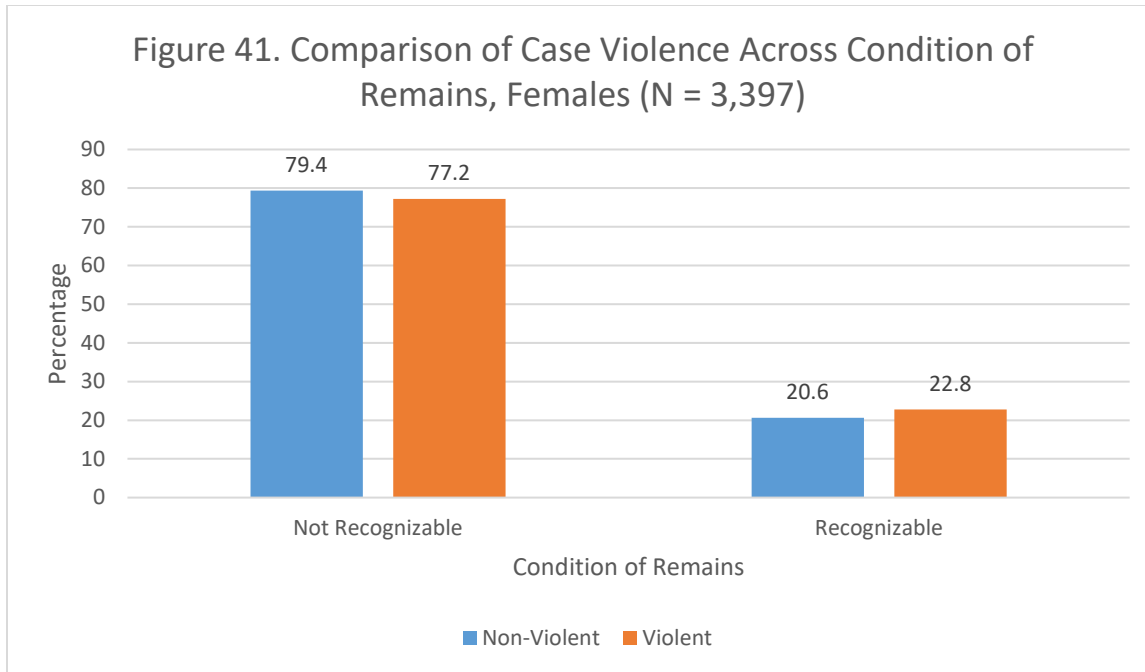
Age Group (N = 13,652)

Among females, 87.9% (n = 1,816) of non-violent cases and 78.0% (n = 653) of violent cases fell into the 21 and over age category (Figure 39). Similarly, among males, 95.6% (n = 8,790) of non-violent cases and 91.2% (n = 1,418) of violent cases fell in the 21 and over age group (Figure 40).



Condition of Remains (N = 15,524)

Among females, 79.4% (n = 1,966) of non-violent cases and 77.2% (n = 712) of violent cases included non-recognizable remains (Figure 41). Similarly, among males, 74.4% (n = 7,793) of non-violent cases and 70.2% (n = 1,162) of violent cases involved remains that were not recognizable (Figure 42).



Age of Case (N = 15,841)*

Among female non-violent cases, 43.2% (n = 1,088) fell into the 20+ year group as did the majority of violent cases (72.7%, n = 684) (Figure 43). Similarly, among male non-violent cases, 41.0% (n = 4,369) fell into the 20+ year age group while 67.4% (n = 1,155) of male violent cases also fell into this same age group (Figure 44).

Figure 43. Comparison of Case Violence By Age of Case Groups, Females (N = 3,458)

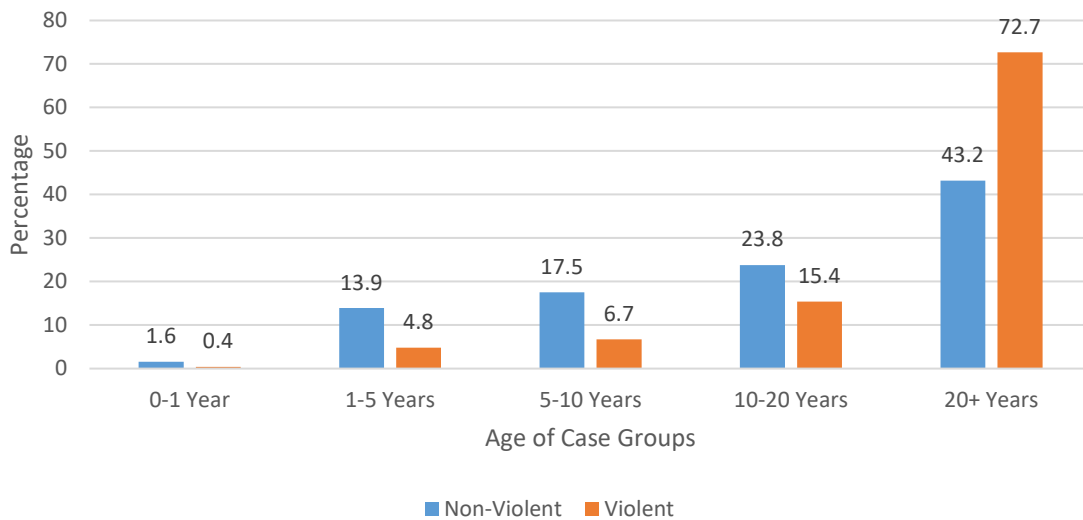
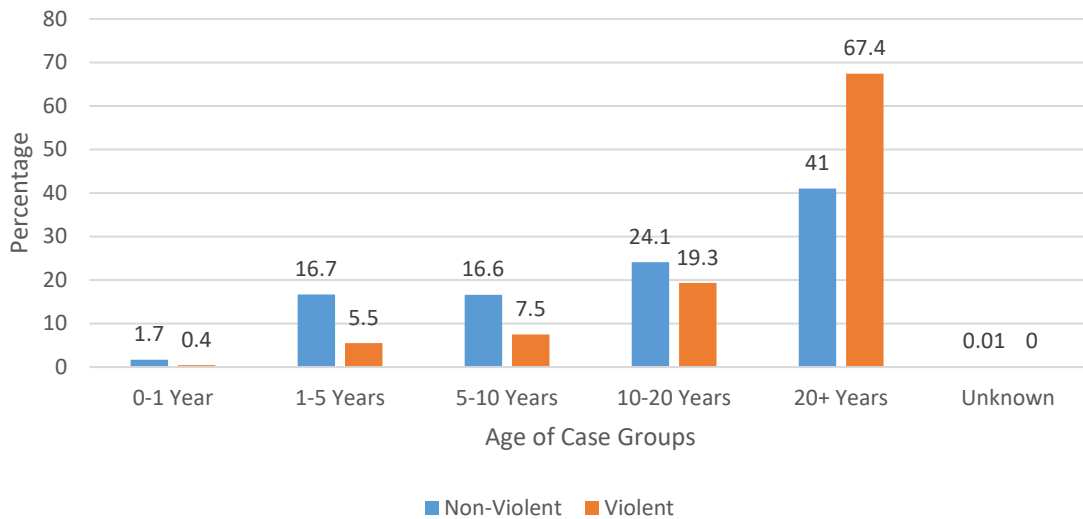


Figure 44. Comparison of Case Violence By Age of Case Groups, Males (N = 12,383)



Resolved Cases (N = 3,605)

NamUs Assisted (N = 2,171)

Among females, 73.2% (n = 300) of non-violent cases and 75.7% (n = 112) of violent cases indicated NamUs assistance (Figure 45). Likewise, 72.2% (n = 1,027) of male non-violent cases and 61.6% (n = 117) of male violent cases also reported NamUs assistance.

Figure 45. Comparison of Case Violence by NamUs Assistance in Case Resolution, Females (N = 558)

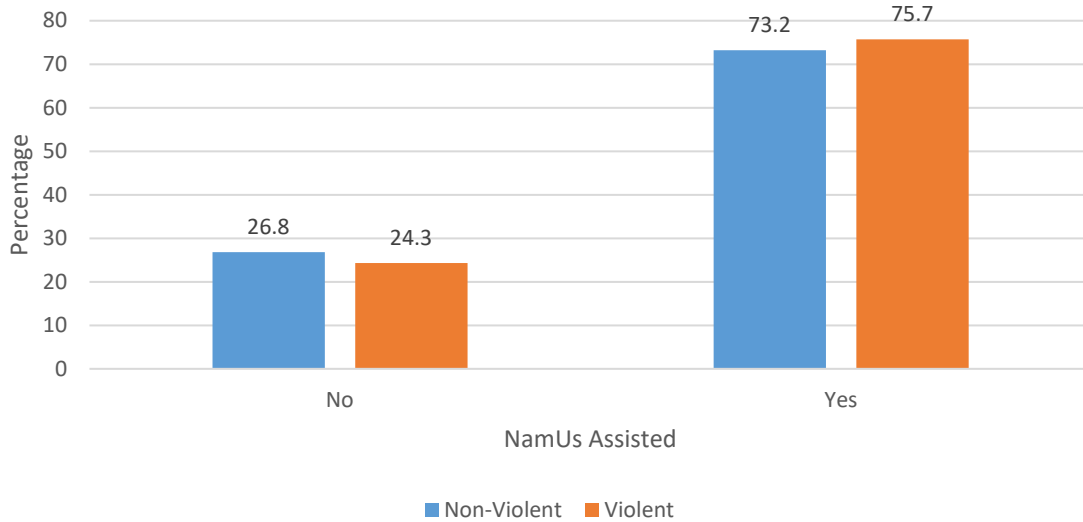
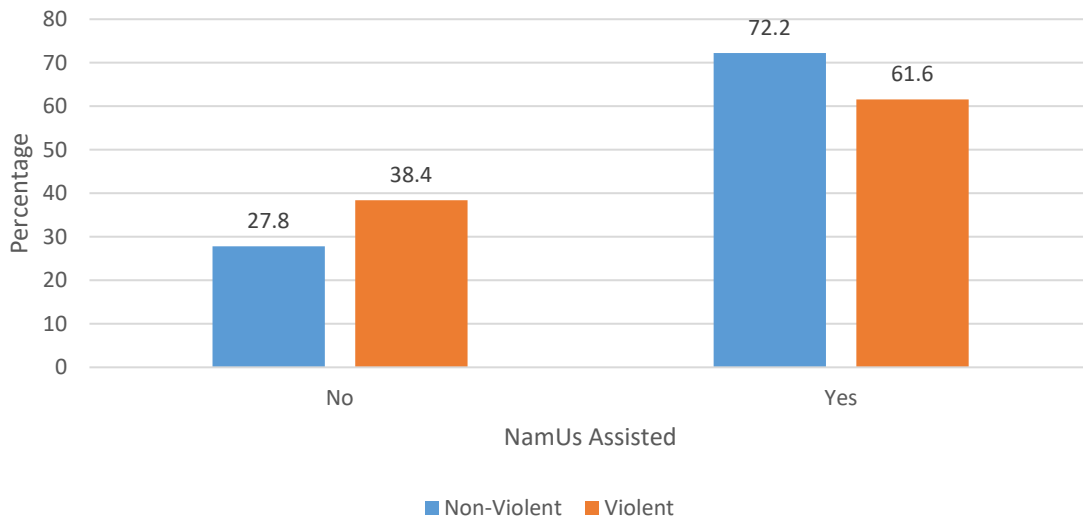
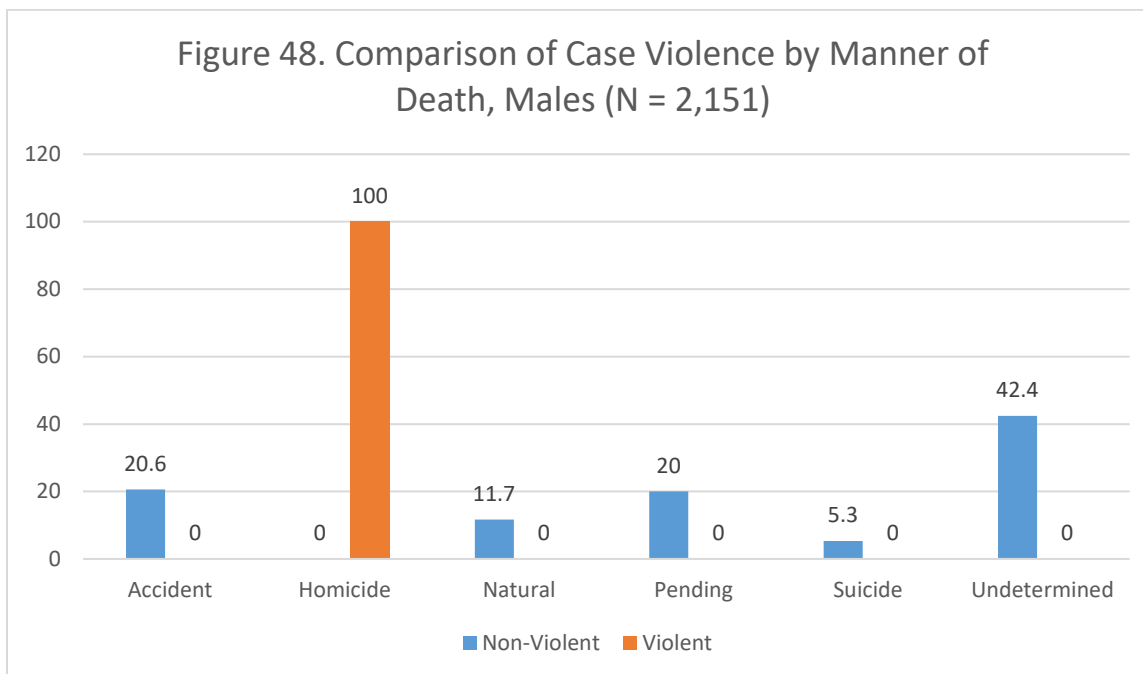
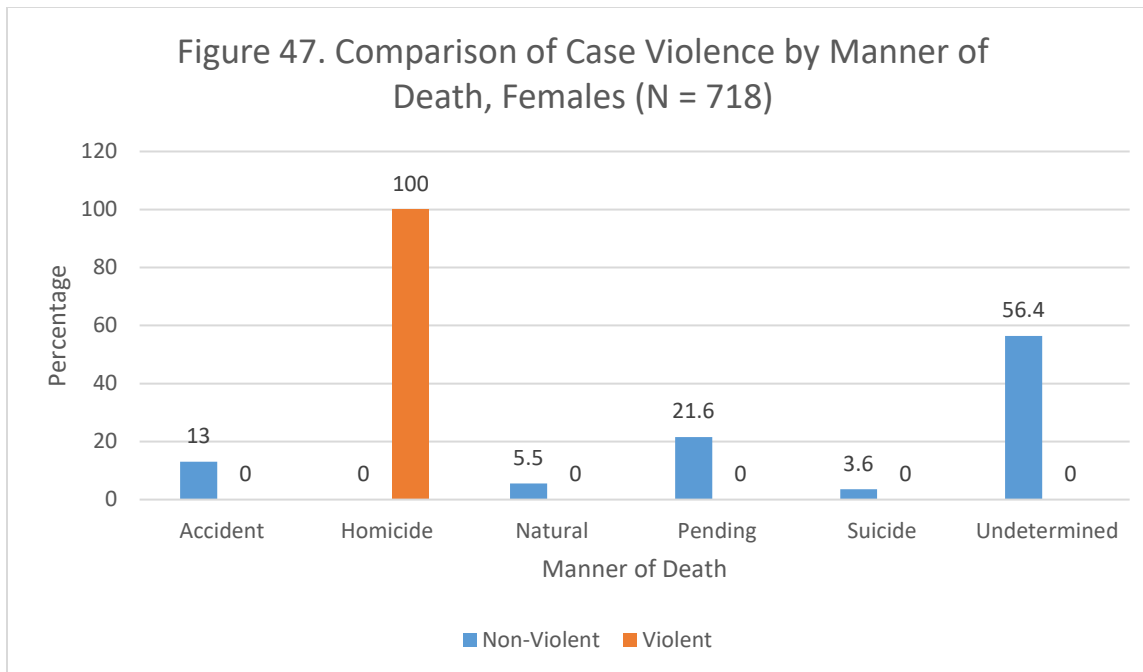


Figure 46. Comparison of Case Violence by NamUs Assistance in Case Resolution, Males (N = 1,613)



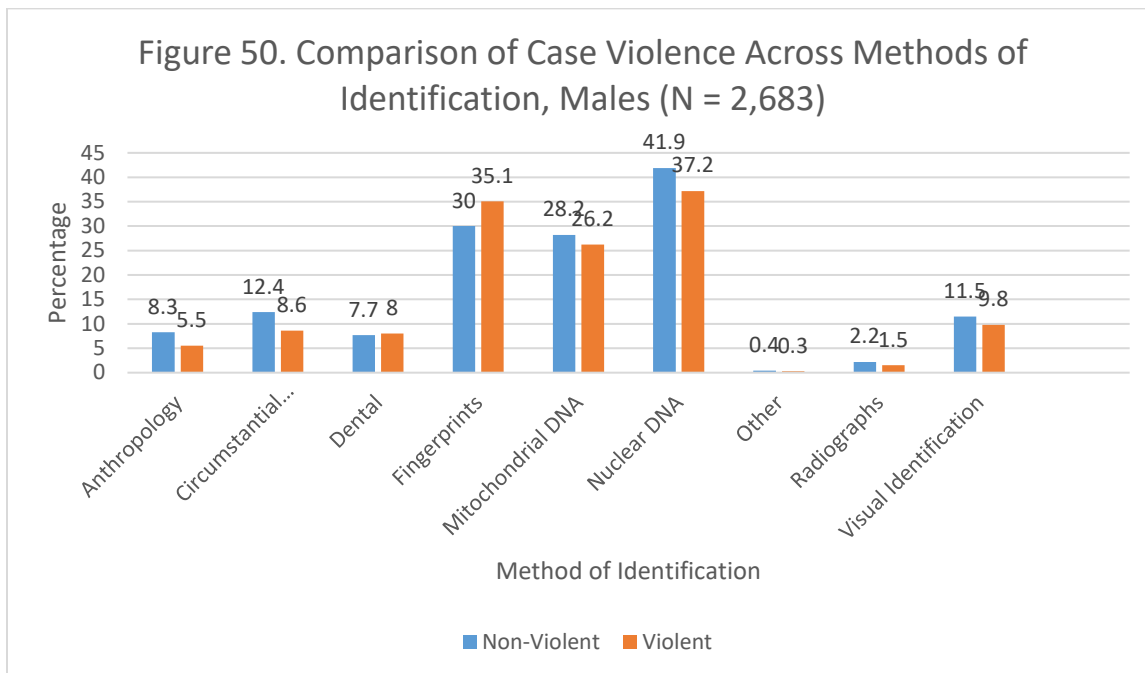
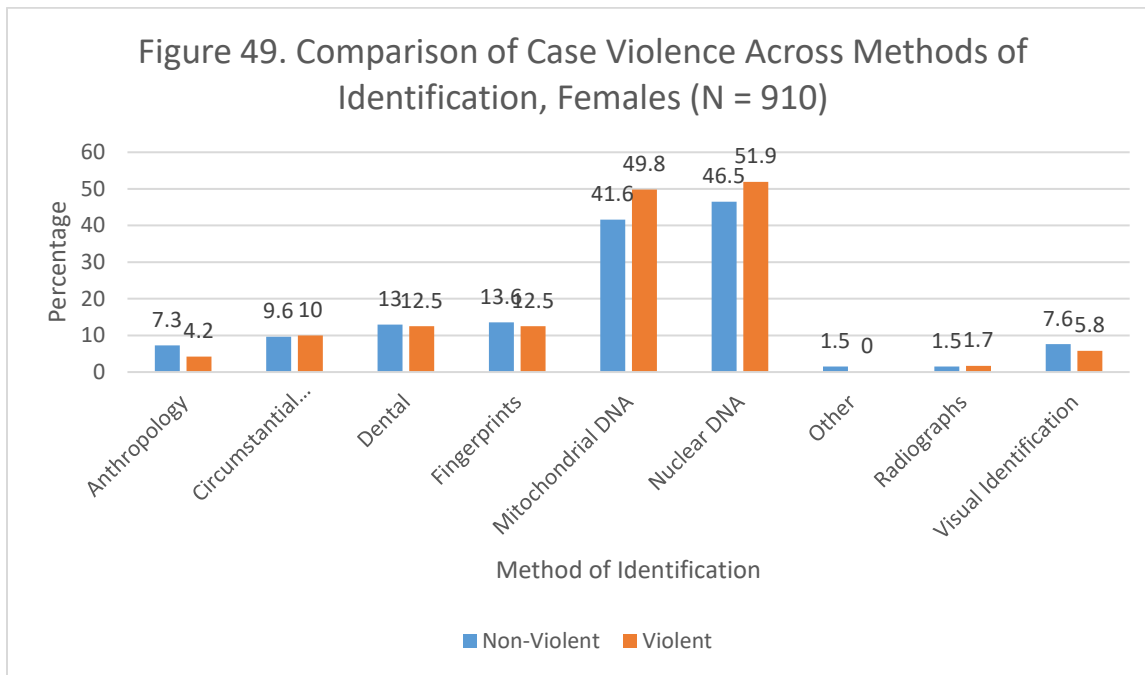
Manner of Death (N = 2,869)*

Overall, all violent cases had a MOD of homicide. Among non-violent cases, the majority of MOD across both genders and overall was undetermined with 56.4% (n = 269) of females (Figure 47) and 42.4% (n = 774) of males (Figure 48) having an undetermined MOD.



Combined Method of ID (N = 3,593)*

Among females, the highest proportion of method of identification was based on short tandem repeat (STR) typing for both non-violent (46.5%, n = 311) and violent (51.9%, n = 125) cases (Figure 49). This trend was also observed among males with 41.9% (n = 987) of non-violent cases and 37.2% (n = 121) of violent cases (Figure 50). However, among females, only 13.6% of non-violent cases and 12.5% of violent cases were resolved using fingerprints while among males, 30% of non-violent cases and 35.1% of violent cases were resolved using fingerprints.

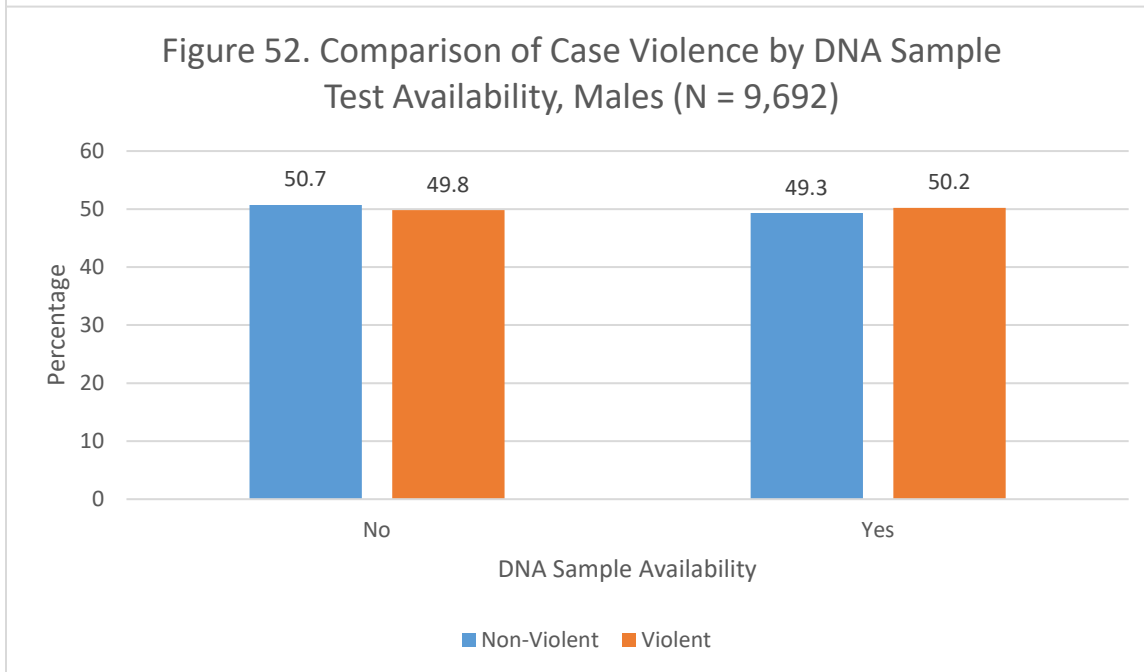
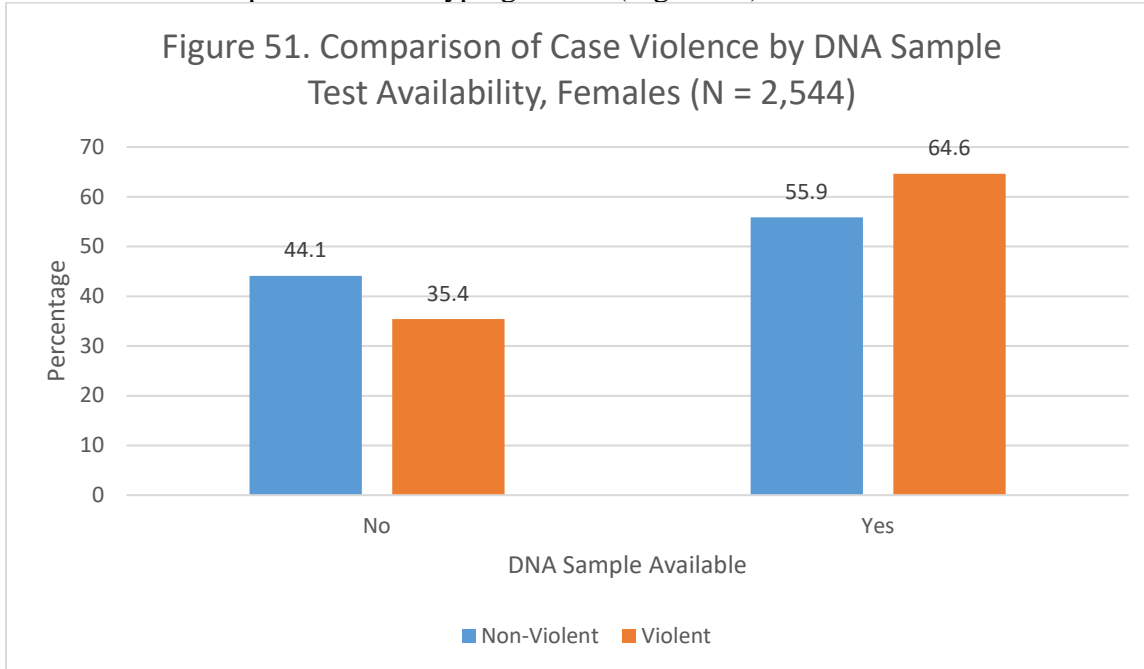


Unresolved Cases (N = 12,236)

DNA Availability (N = 12,236)

Overall, as shown in Table 17, most cases had at least one DNA reference sample uploaded and tested. Among females, 55.9% (n = 1,031) of non-violent cases and 64.6% (n = 452) of violent cases had DNA testing information for at least one sample uploaded (Figure 51). Among males,

49.3% (n = 4,091) of non-violent cases had DNA typing uploaded while 50.2% (n = 697) of violent cases had uploaded DNA typing results (Figure 52).



Recommendations

Based on the case classification and analysis, the following recommendations are made. The recommendations are put forth in hopes of improving the identification of cases involving violence, as well as enhancing the investigation and resolution of such cases based on the unique characteristics of violent cases that were uncovered. Further, these recommendations serve to enhance data collection, which will serve to improve future research, evaluation, and analysis using NamUs case data. In all, these recommendations further the NamUs mission by bridging improvements in data and information gathering, enhancements in database technology, and NamUs' human services to better resolve cases involving violence.

As shown in Figures 41 and 42, the vast majority of both violent and non-violent cases for both males and females had unrecognizable remains.

- **Recommendation 1:** Given the high levels of unrecognizable remains, ensure that biometric characteristics, such as DNA, fingerprints, and dental records, are obtained to assist in the case resolution.
- **Recommendation 2:** For cases with unrecognizable remains, NamUs should develop forensic art resources to create composite sketches or enhance post-mortem images to assist with public identification of such cases.
- **Recommendation 3:** Implement a quality assurance process for resolved cases to verify that the initial entries for population/ethnicity, age, gender, etc. are indeed correct in the case file prior to case closeout.

As shown in Figures 43 and 44, the age of the case for the majority of both violent and non-violent cases for both males and females was 20 years or more.

- **Recommendation 4:** Such cases are more likely to be remains that were buried, and as a major barrier to exhumation and subsequent identification is the financial barrier to law enforcement, NamUs should provide funding to exhume such remains for further identification purposes.

As shown in Figures 47 and 48, the majority of both male (42.4%) and female (56.4%) non-violent cases had an undetermined MOD.

- **Recommendation 5:** Implement a quality assurance process for resolved cases to verify MOD, when possible, and to ensure the case file is complete and up-to-date with information uncovered through the resolution.
- **Recommendation 6:** For cases in which the ME/C was unable to enter MOD into NamUs due to confidentiality or due to ongoing investigation, provide the opportunity for such personnel to enter such information and make it visible only to NamUs personnel and the associated criminal justice agency.
- **Recommendation 7:** Partnerships should be explored between NamUs and the agencies that hold the death certificates to cross-check NamUs MOD against official death certificates held by outside agencies.

As shown in Figures 49 and 50, a high proportion of both male and female violent and non-violent cases were resolved using some type of DNA testing; however, the proportion of male violent and non-violent cases that were resolved using fingerprints was over twice the proportion of violent and non-violent female cases.

- **Recommendation 8:** Beginning with unidentified males, specific effort should be made to obtain fingerprint records for comparison to fingerprints systems, as NamUs fingerprint service is more likely to get a hit on male cases because they are more represented in fingerprint databases.

As shown in Figures 51 and 52, a higher proportion of unresolved, active female violent cases have a DNA sample available (64.6%) compared to male violent cases (50.2%).

- **Recommendation 9:** Encourage RPSs to work with case owners for all violent cases in order to obtain DNA samples.

•

Conclusion

This investigation into UP case violence and the case characteristics associated with such violence within the NamUs database offers insights into the differences that exist between violent and non-violent cases. In doing so, the results enable the making of recommendations for systems improvement and expansion to better understand case violence, which will enhance the ability for NamUs to respond to and assist with UP cases involving violence by working with law enforcement to obtain DNA samples (mitochondrial and nuclear), fingerprints, or to assist in exhumations that may be financially infeasible for law enforcement agencies, for example. Further, the findings of this project highlight areas for future research and evaluation to fill critical gaps in knowledge that could not be answered through this study due to limitations of the project. Overall, however, the results presented in this report provide more detailed information that provide direction for the improvement and expansion of the NamUs system, including the database and the various service divisions that make-up the NamUs program, such as the inclusion and mandating of various data fields to better capture violence in NamUs, and increased efforts to obtain biometric samples for identification of cases.

Limitations

The UP classification and analysis had several limitations. As with the MP classification, operational data were used for systems improvement purposes. The data were never intended to be used for such purposes. They represent crowd sourced data as opposed to data collected in a systematic manner such as data collected using survey data collection techniques. As such, they may be deficient for a systems improvement study.

Second, the classification dictionary may have been deficient. Coverage of all relevant keywords may not have been achieved, including the coverage of all possible inflected forms of each keyword. Due to these likely deficiencies, violent cases may have been under-classified.

Further, only 20% of cases were used for secondary coding of the computer classifications, and only 10% of the 20% of second coded cases were used for tertiary coding. These levels may not have been adequate to accurately assess and verify the dictionary's validity. Thus, all deficiencies within the dictionary may not have been ascertained using these numbers of cases.

Additionally, inter-rater agreement was calculated as percent agreement between the computer and a human coder, and subsequently, between two human coders. While the percent agreements were near or above 90%, this threshold may not have been high enough to accept the results in their totality as completely valid. As such, the final classifications of all cases may not have been acceptable or appropriate due to misclassifications accepted as true and valid based on the inter-rater agreement thresholds.

Finally, a large proportion of cases had remains that were not identifiable, which likely impacted the accuracy and precision of case file information, particularly demographic information. In particular, cases in which the entire body was not recovered or in which there was severe decomposition may have been more greatly affected by this issue. In such cases, characteristics such as population/ethnicity, gender, age, time since death, or other characteristics may have been difficult if not impossible to estimate.

IV. Special Section: Missing and Unidentified Indigenous Persons

Background

The serious issue of missing and murdered indigenous persons (MMIP), particularly women and girls, has gained national attention through the grassroots advocacy and work of tribes, tribal organizations, and tribal members (for purposes of this report, “indigenous persons” refers to American Indian and Alaska Native [AI/AN] persons). This advocacy has resulted in increased efforts to continue raising awareness of and enhance response to the issues as exemplified by the establishment of The Presidential Task Force on Missing and Murdered American Indians and Alaska Natives known as Operation Lady Justice (OLJ, n.d.). The purpose of this Task Force is to engage in consultation with tribes and villages; to work to enhance response to missing or murdered AI/AN persons through the development of model protocols for law enforcement response, and data sharing and database use; to establish multi-disciplinary cold case review teams; and to design an educational outreach and awareness campaign (OLJ, n.d.). The establishment of this Task Force, including its mandates, exemplifies the broad scope and pressing nature of these issues.

Violence among AI/AN Persons

As reported by Rosay (2016), AI/AN persons, both women and men, face high levels of violence victimization. Overall, in a nationally-representative sample of self-identified AI/AN women and men, 84.3% of women and 81.6% of men had ever experienced violence in their lifetime. These rates are higher than for non-Hispanic Whites. AI/AN women experienced 1.2 times and AI/AN men experienced 1.3 times the rates of victimization as non-Hispanic Whites. More specifically, 66.4% of women and 73.0% of men have experienced psychological aggression by an intimate partner, 55.5% of women and 43.2% of men have experienced physical violence by an intimate partner, 56.1% of women and 27.5% of men have experienced sexual violence, and 48.8% of women and 18.5% of men have experienced stalking.

Data Issues with MMIP

Tribal advocates have stressed that there are little data to fully understand the true scope and nature of MMIP (Sovereign Bodies Institute [SBI], 2019), particularly in urban areas (Urban Indian Health Institute [UIHI], 2018). While it is unclear exactly how many MMIP there may be, advocates maintain that most databases do not fully capture the true extent of these issues due to underreporting, reports not taken by law enforcement, and misclassification of indigenous persons as non-indigenous, among other reasons. (SBI, 2020; UIHI, 2018). This lack of data completeness and quality has resulted in efforts to improve data collection, including efforts at the policy level. For example, the federal Savanna’s Act mandates improved coordination of, access to, and reporting of statistics on MMIP across federal, state, and tribal jurisdictions. Further, it mandates clearer protocols on how such data collection may be improved. The Act includes NamUs as a party to the legislation as one database that should be engaged in data collection processes and process improvements (Congress.gov, 2020 Oct 10).

Various entities or organizations are making attempts to improve data collection and quality regarding MMIP. Among non-governmental organizations, SBI has worked to create a database tailored specifically to the data issues of indigenous persons and communities that may not be collected by databases for the general population (SBI, 2019). UIHI has also engaged in data collection from law enforcement agencies in urban population centers to better describe the issues among urban AI/AN persons (UIHI, 2018). Overall, efforts from community and non-governmental organizations have been at the forefront of discussion regarding the MMIP issues.

Among governmental systems, NamUs has actively worked with the Bureau of Indian Affairs at the U.S. Department of the Interior to define and implement data fields specifically for AI/AN persons. These data fields include tribal affiliation/enrollment, primary residence on tribal land, and missing from or found on tribal land, among others. Additionally, NamUs has worked to increase its outreach to tribal communities through trainings, webinars, and monthly publication of tribal-specific data informational sheets.

Methods

Case Classification

Case classification procedures to determine if a case was violent or non-violent have been described above and followed the same procedures outlined in the MP and UP methods sections, respectively.

Case Analysis

Variables

Variables used for this special section included the same variables as used in the overall MP and UP sections of the report with the notable exception of population/ethnicity.

Population/ethnicity is not included as all cases involve persons identified as AI/AN alone or in combination with another population/ethnicity. Further, initial attempts were made to use the tribal-specific variables described above; however, for both female and male cases, these variables suffered from high levels of cell counts under 10 cases. As such, it was deemed inappropriate to use these variables, as the results may have been misleading or overreaching what could actually be concluded by data from so few cases.

Analytic Approach

Cases were analyzed using similar methods used in the MP and UP overall analyses.

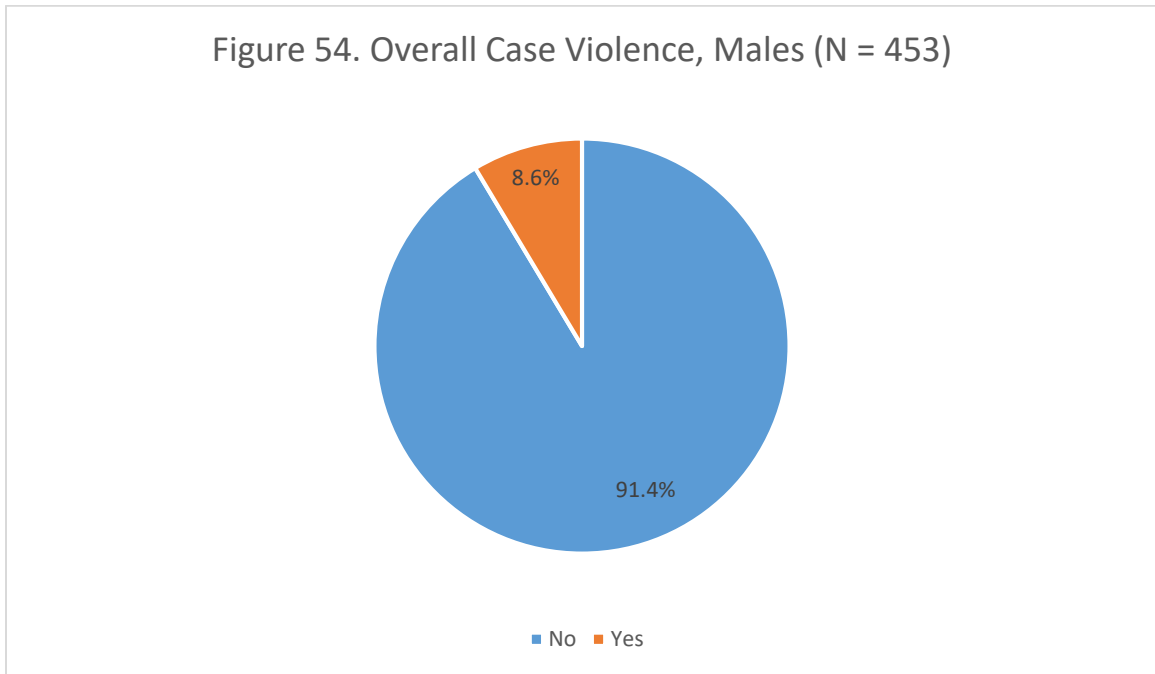
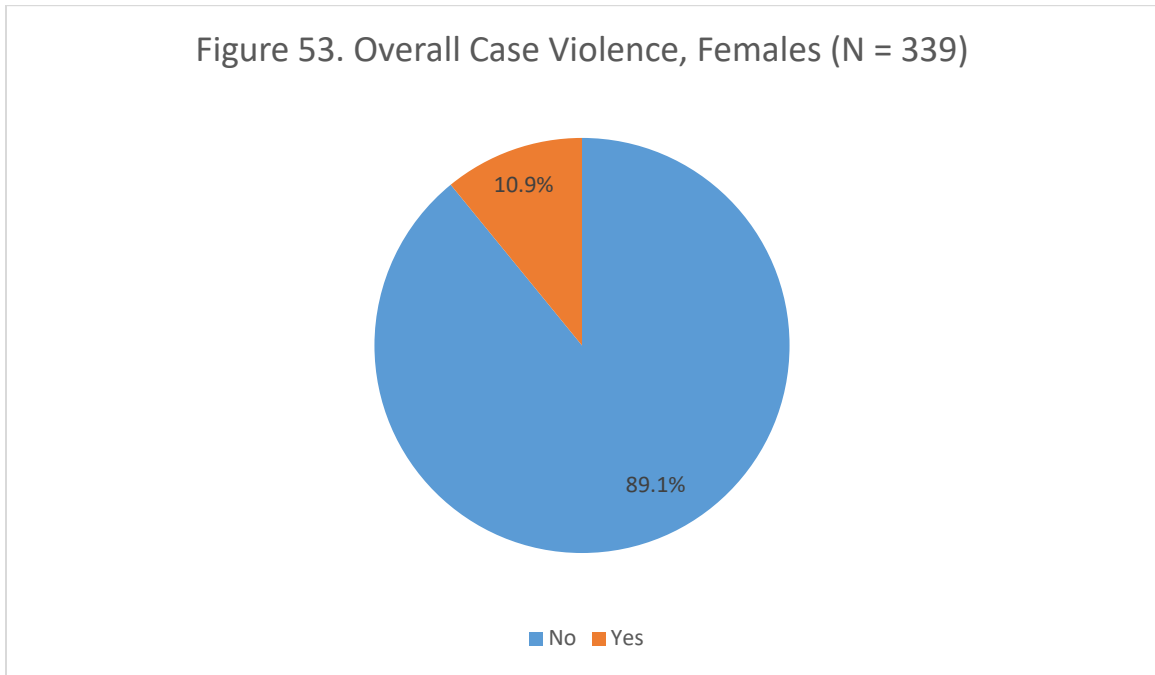
Results: MP

Overall, there were 792 valid cases for classification and analysis. Females made up 42.8% (N = 339) and males encompassed 57.2% (N = 453) of all cases. Figures 53-80 display results for MP cases, including gender stratified findings for types of violence and gender stratified bivariate results comparing case characteristics by presence of case violence across case resolution status. An asterisk (*) next to a variable indicates that there was at least one cell for which the case count was less than 10.

All Cases

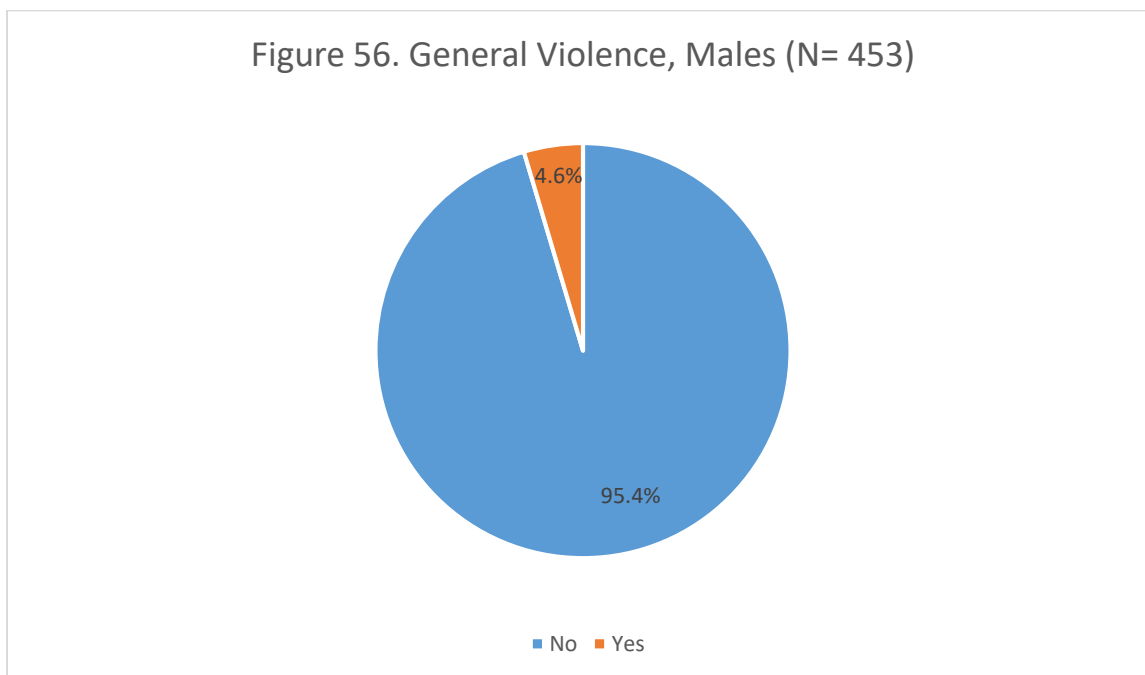
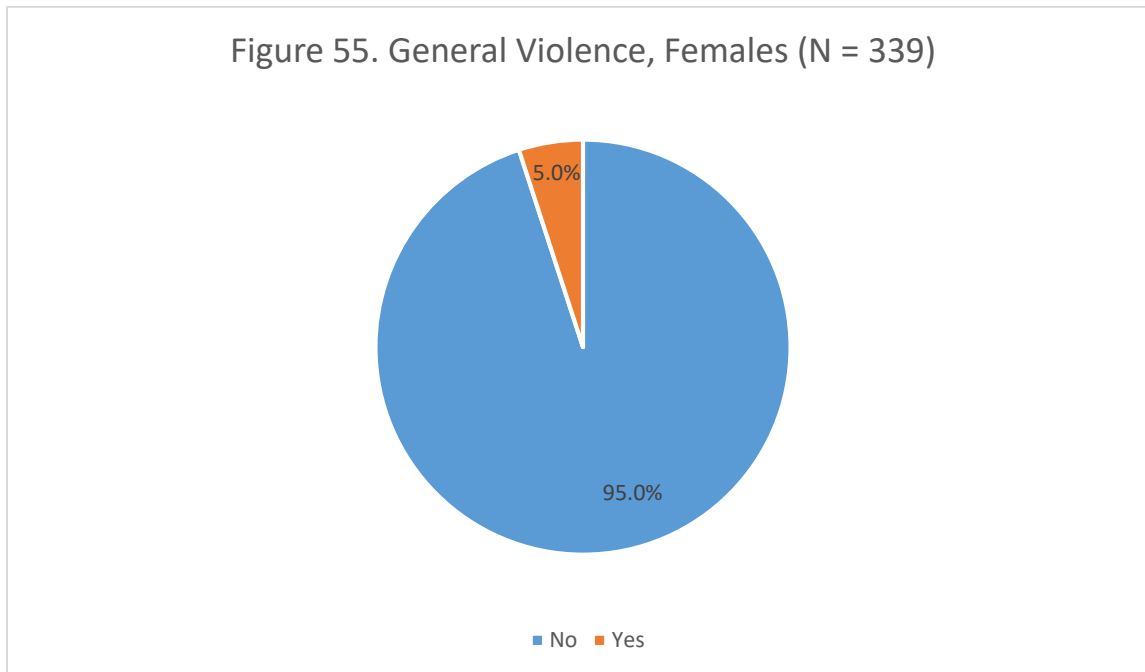
Overall Violence (N = 792)

Among females, 10.9% (n = 37) experienced some form of violence (Figure 53), while 8.6% (n = 39) of males experienced some form of violence (Figure 54).



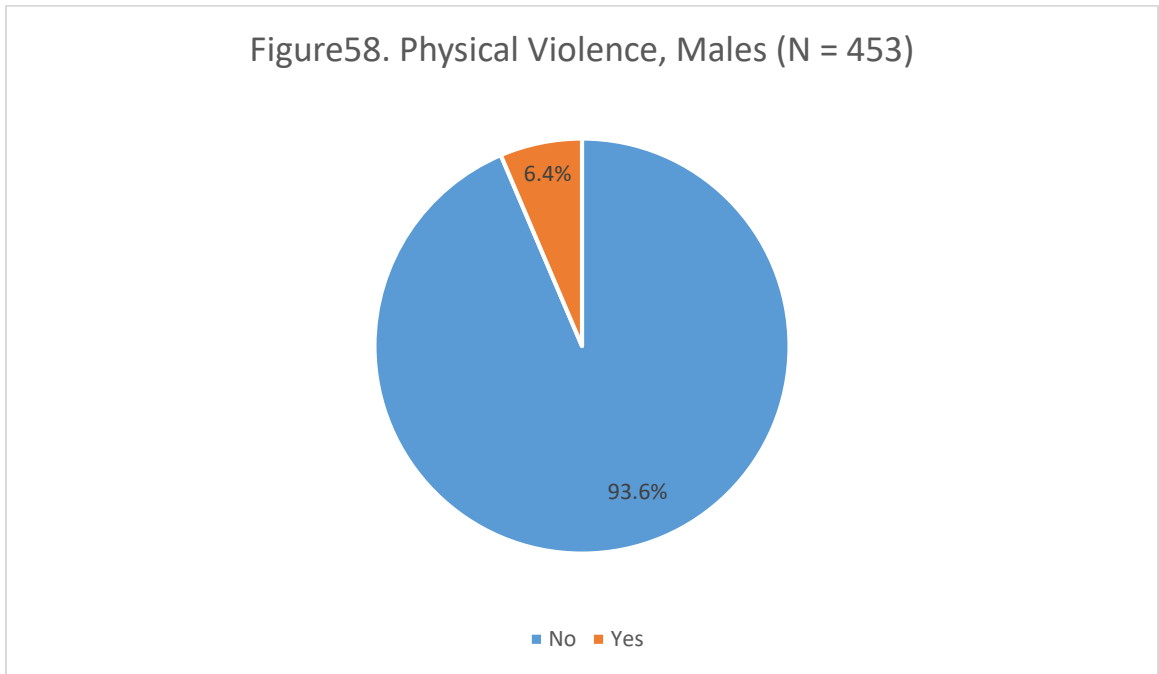
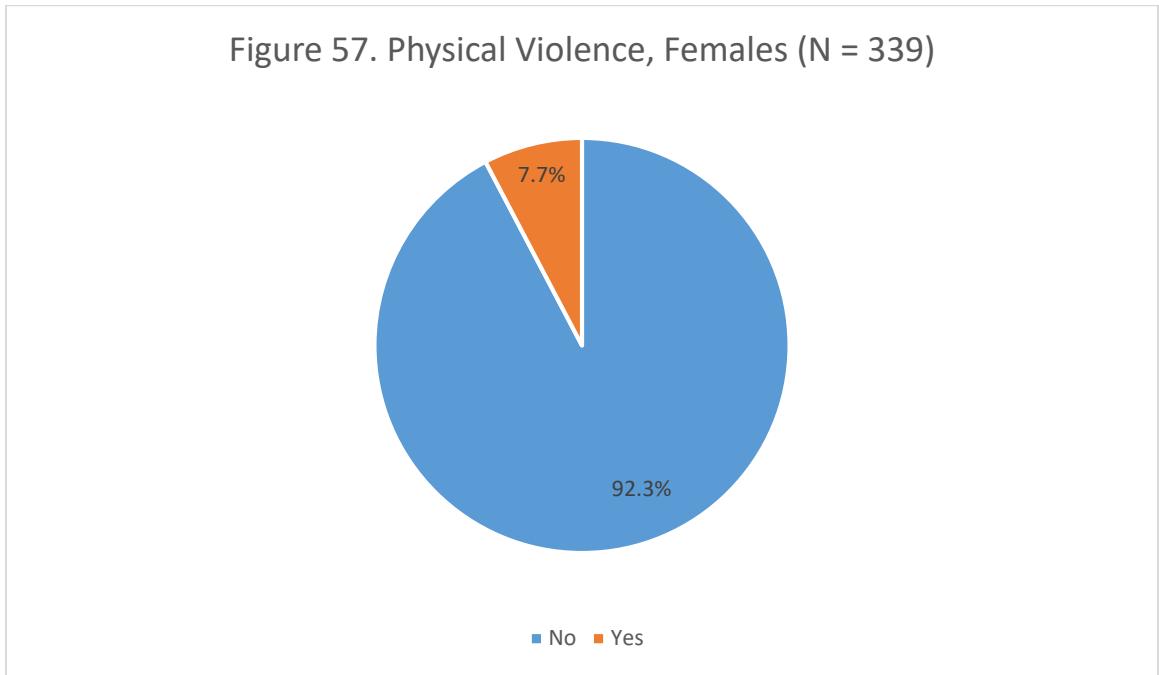
General Violence (N = 792)

As with overall violence, a higher proportion of females (5.0%, n = 17) (Figure 55) than males (4.6%, n = 21) (Figure 56) experienced general violence.



Physical Violence (N = 792)

Among females , 7.7% (26) experienced physical violence (Figure 57), while 6.4% (n = 29) of males experienced physical violence (Figure 58).

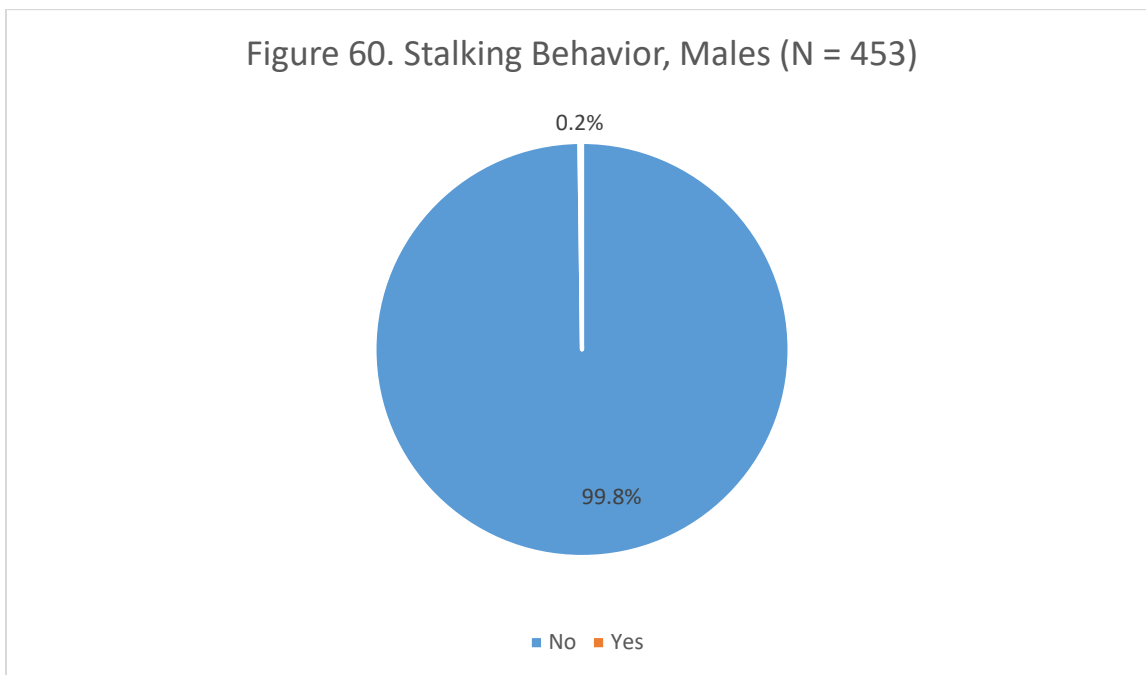
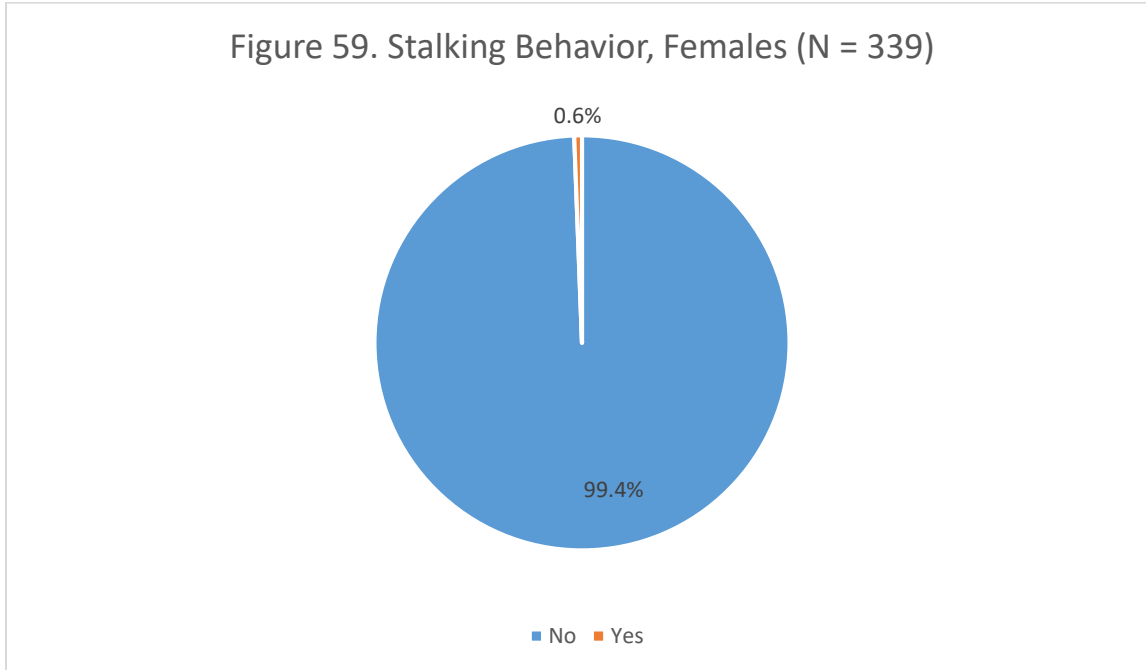


Sexual Violence (N = 792)*

No cases for females or for males involved sexual violence.

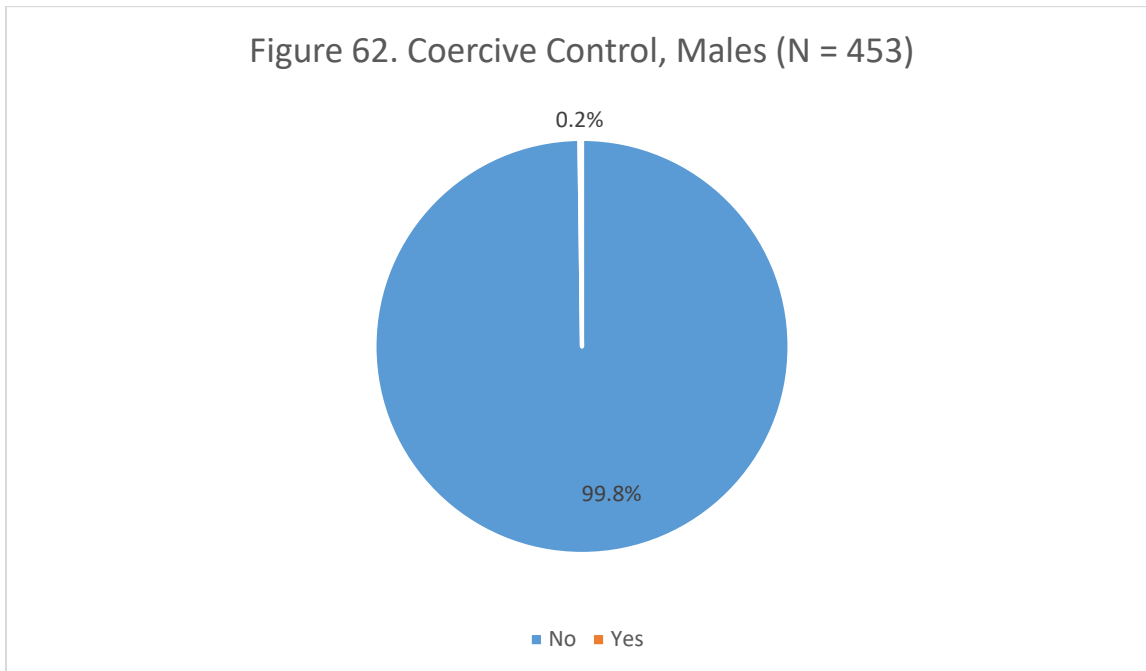
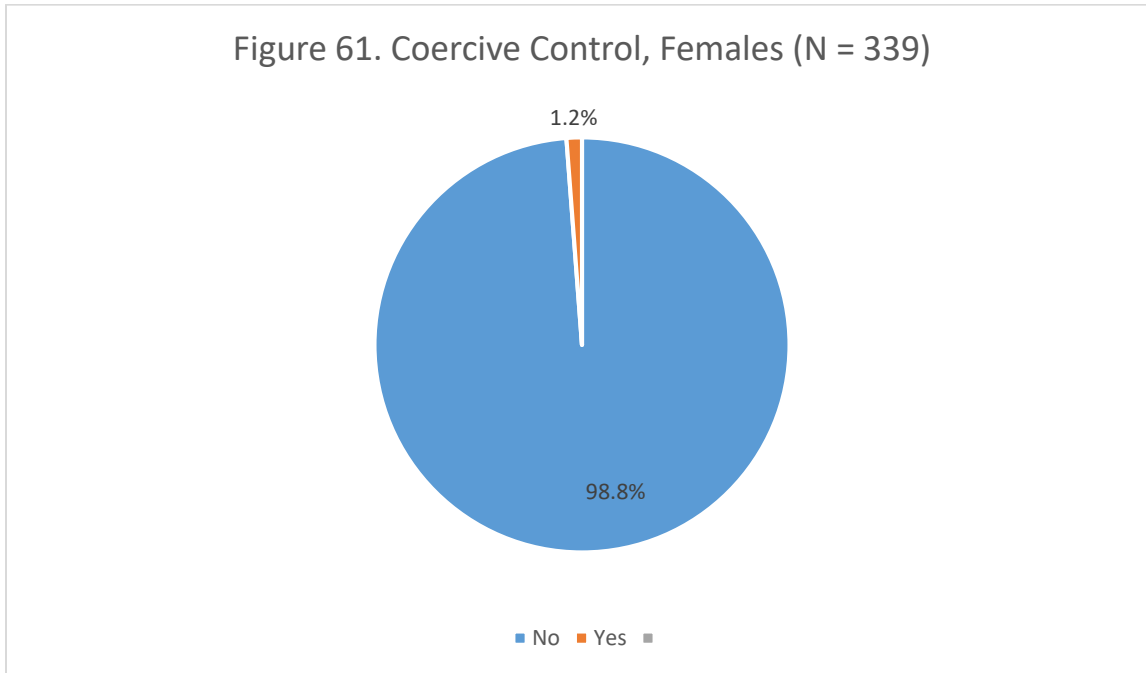
Stalking Behavior (N = 792)*

For females, 0.6% (n = 2) experienced some type of stalking behavior (Figure 59), while 0.2% (n = 1) of males experienced stalking behavior (Figure 60).



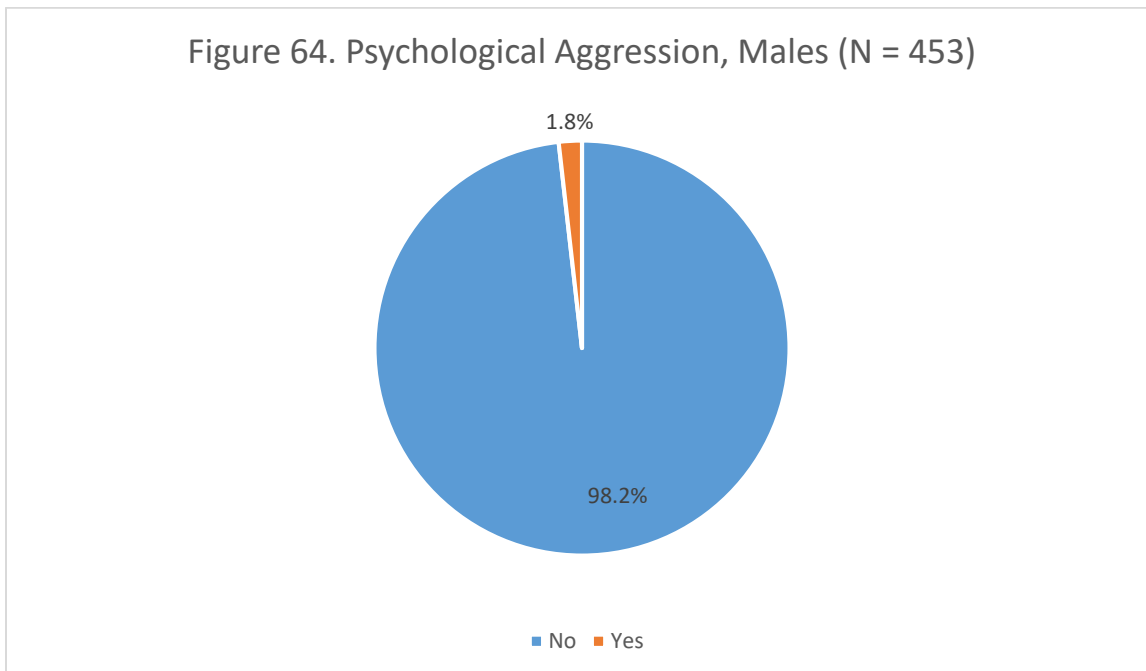
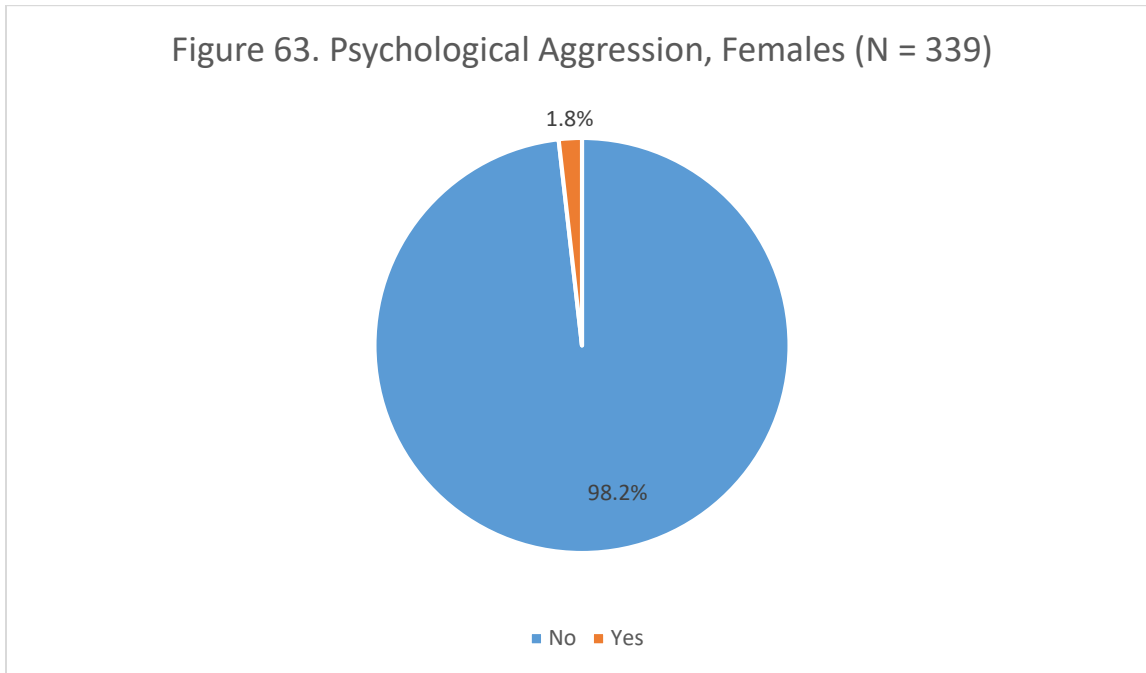
Coercive Control (N = 792)*

Among females, 1.2% (n = 4) experienced this type of violence (Figure 61), while 0.2% (n = 1) of males did as well (Figure 62).



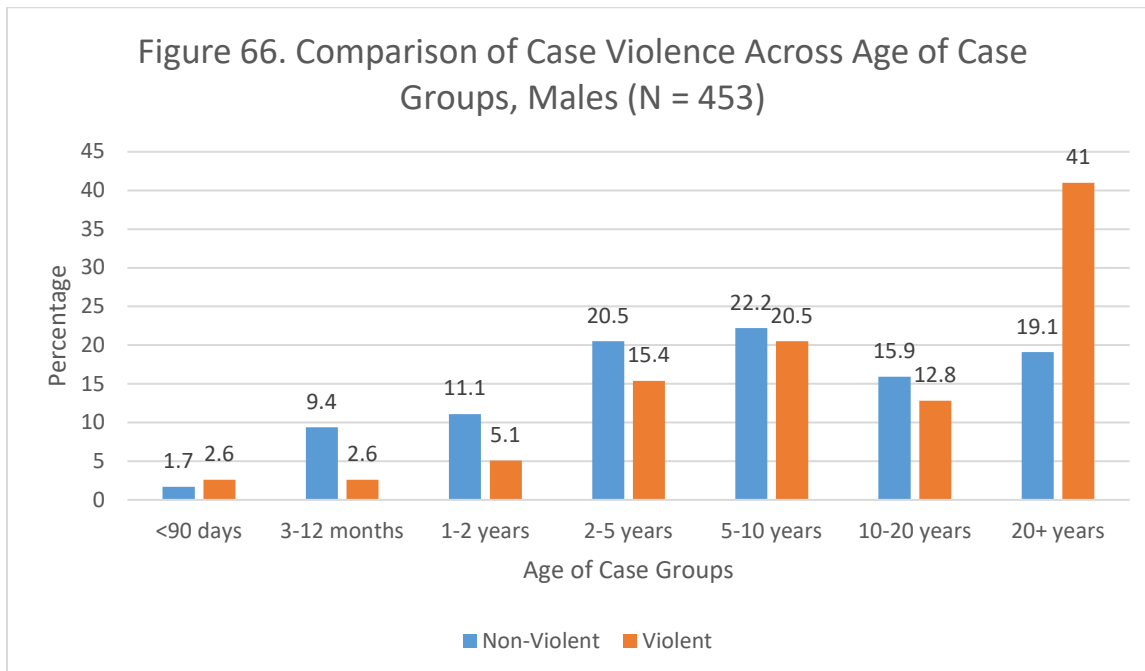
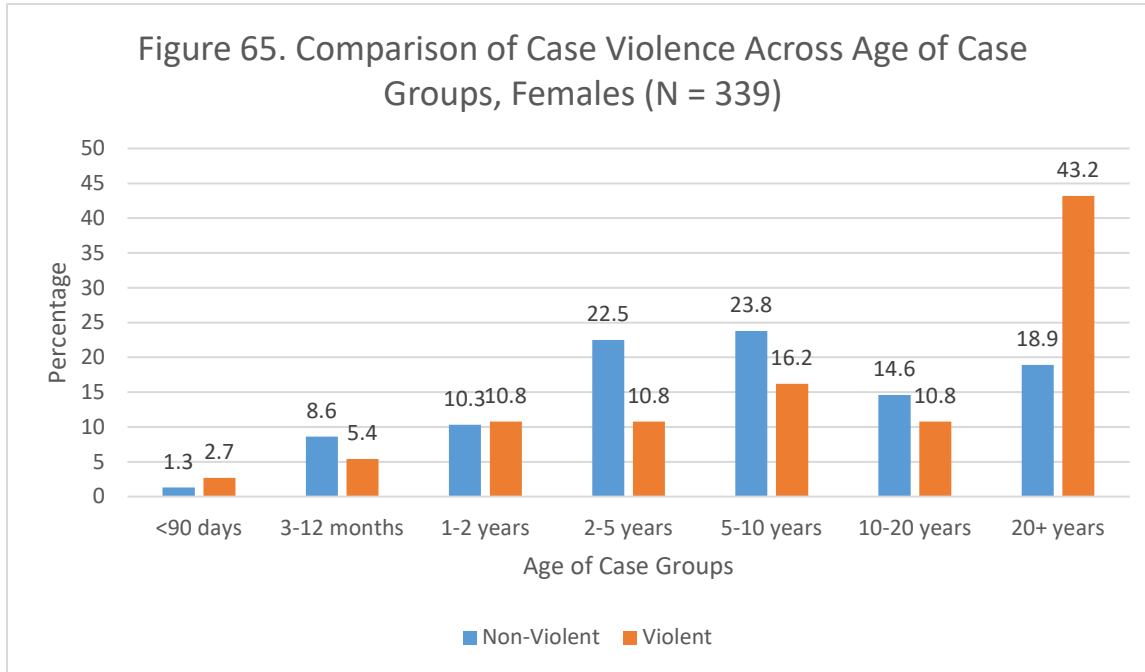
Psychological Aggression (N = 792)*

Overall, 1.8% (n = 6) of females (Figure 63) and 1.8% (n = 8) of males (Figure 64) experienced psychological aggression.



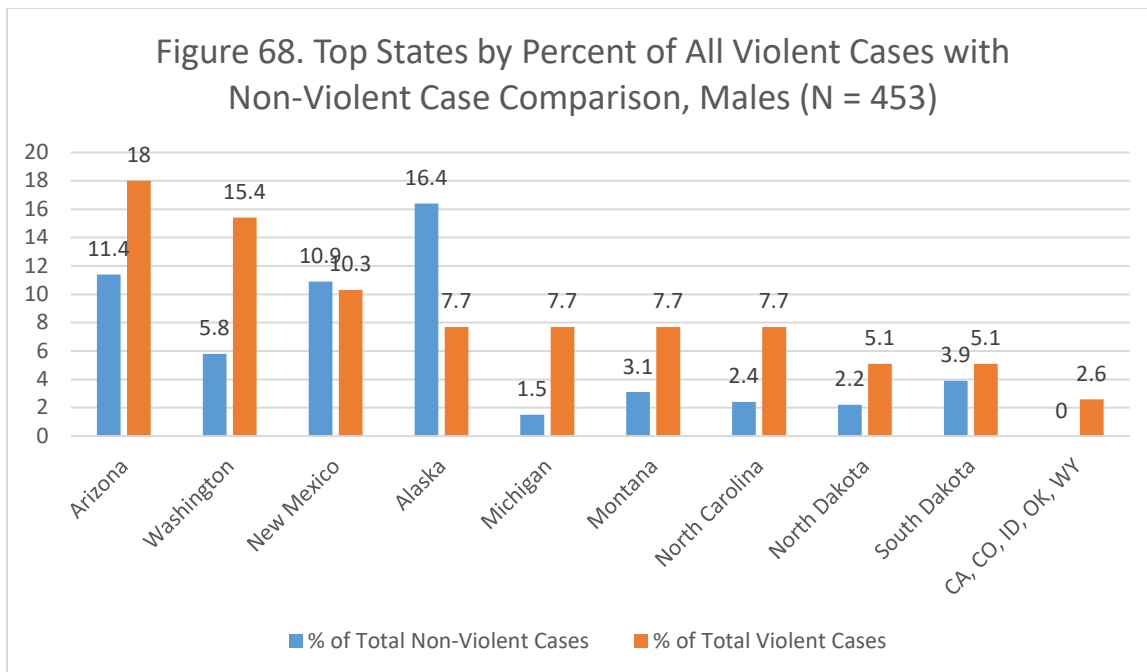
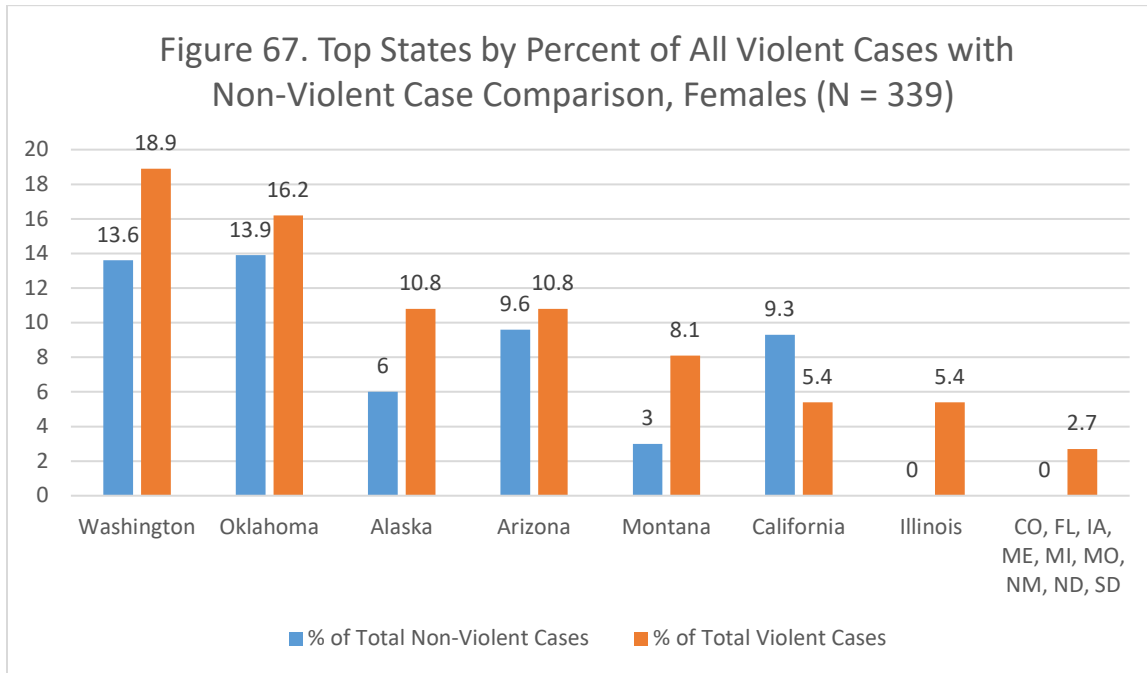
Age of Case (Groups) (N = 792)*

Among female non-violent cases, the highest proportion of cases were 5-10 years old (23.8%, n = 72) while the highest proportion of violent cases were 20+ years old (43.2%, n = 16) (Figure 65). This trend was mirrored among male non-violent cases where the highest proportion of cases were observed among 5-10 year old cases (22.2%, n = 92) while the highest proportion of violent cases were 20+ years old (41.0%, n = 16) (Figure 66).



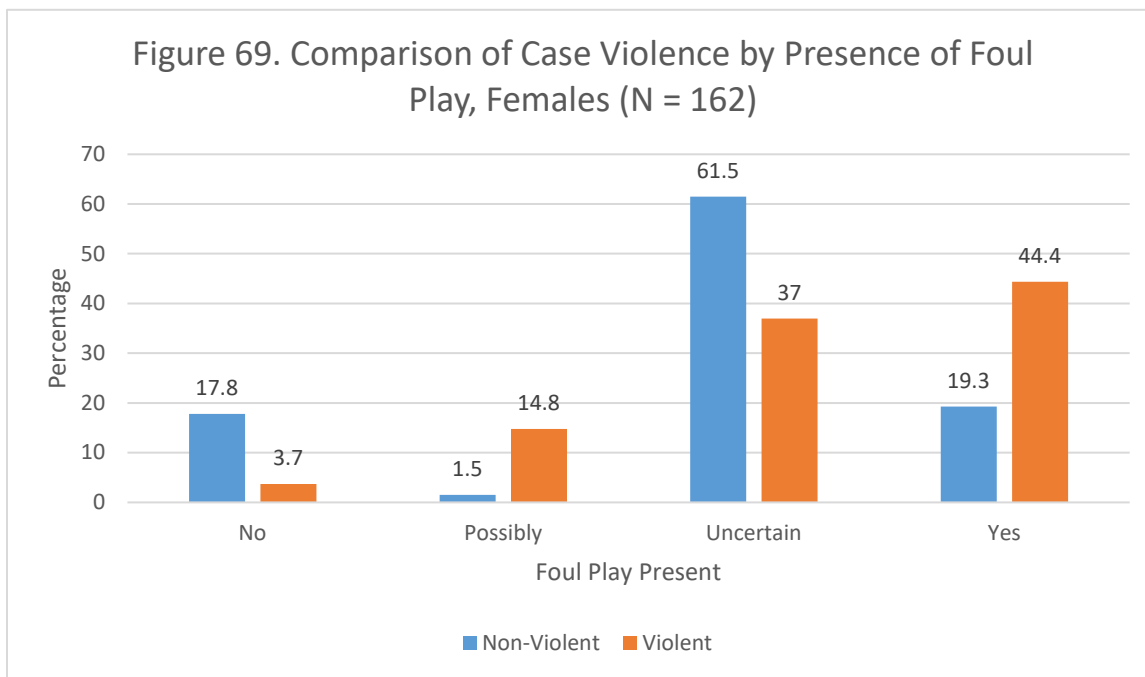
State Name (N = 792)*

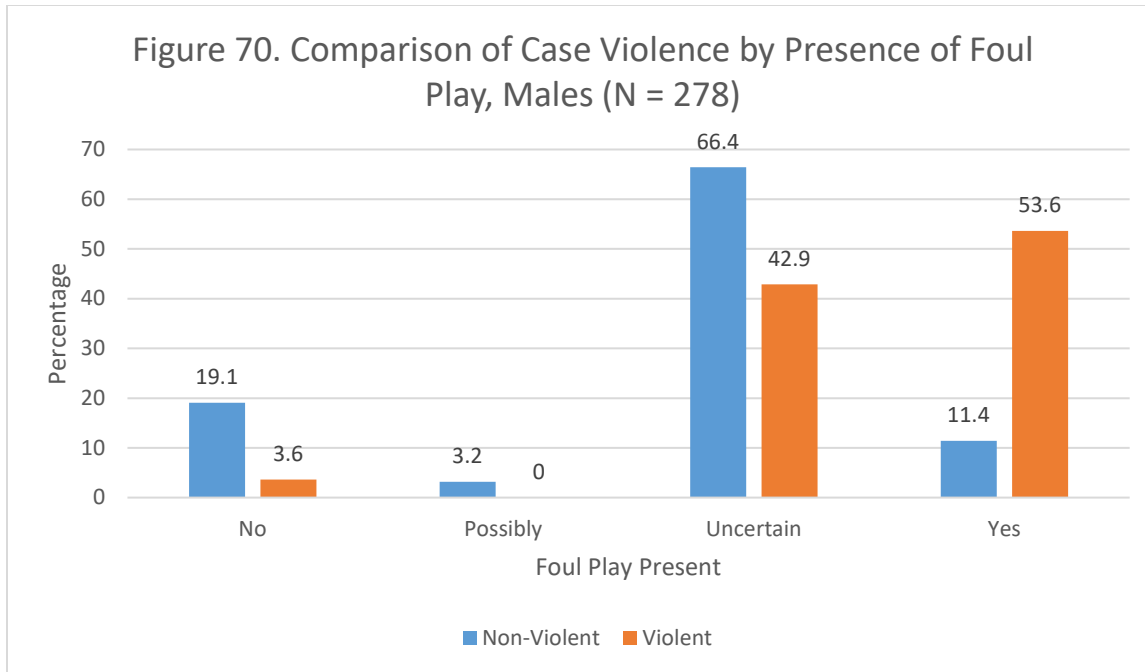
Figures 67 and 68 show the states representing the highest proportion of violent cases in NamUs with their non-violent case percentages shown for comparison. Among females, Washington state had the highest proportion of violent cases, representing 18.9% (n = 7) of all violent cases (Figure 67). Among males, Arizona had the highest proportion of violent cases, representing 18.0% (n = 7) of all violent cases (Figure 68). The numbers presented represent absolute numbers of case violence, and do not represent rates of violence within the states; as such, the numbers may reflect higher usage of the NamUs system as opposed to actual higher levels of violent cases within these states.



Foul Play (N = 410)*

Among female non-violent cases, 61.5% (n = 83) of cases indicated uncertainty if foul play was likely (Figure 69). However, among female violent cases, 44.4% (n = 12) of cases indicated likely foul play. A high proportion of cases not classified as violent did indeed have the foul play indicator checked (19.3%, n = 26). Among male non-violent cases, 66.4% (n = 146) indicated uncertainty as to the presence of foul play (Figure 70). Among male violent cases, 53.6% (n = 15) indicated likely foul play. Again, as with female results, a high proportion of cases that were not classified in the classification process as being violent had that foul play indicator checked (11.4%, n = 25).





Age Missing (Group) (N = 792)*

Among females, the highest percentage of cases were for the under 18 age group for both non-violent (34.8%, n = 105) and violent (27.0%, n = 10) categories (Figure 71). Among males, the highest proportion of non-violent cases was for the 21-30 age group (26.8%, n = 111), and the highest proportion of violent cases was for the same age group (30.8%, n = 12) (Figure 72).

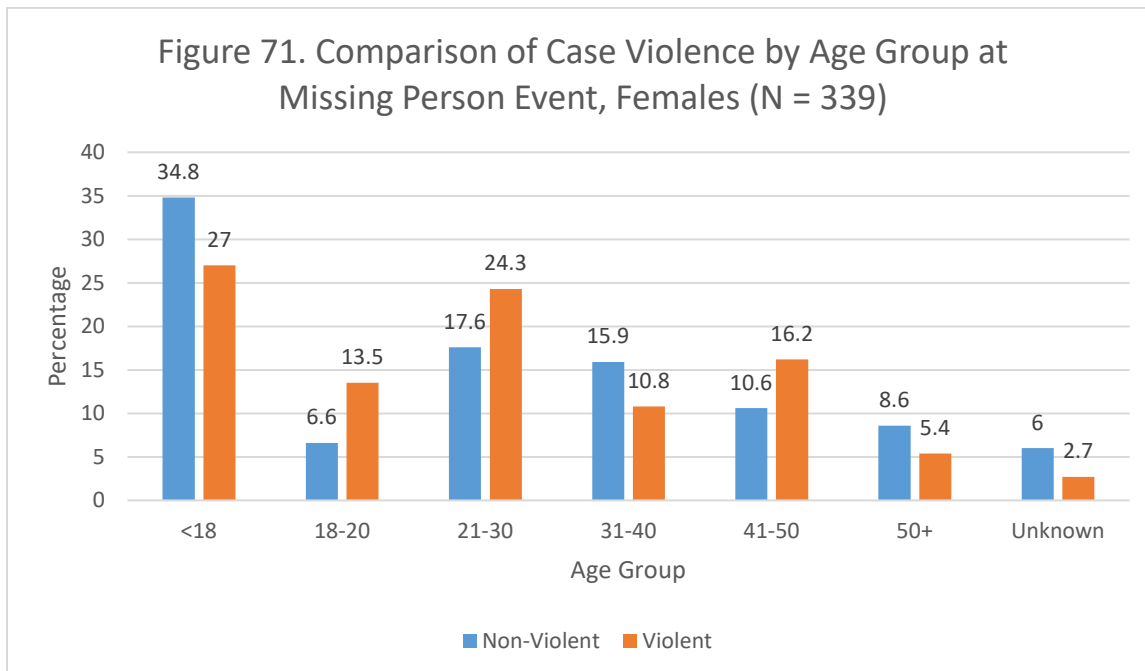
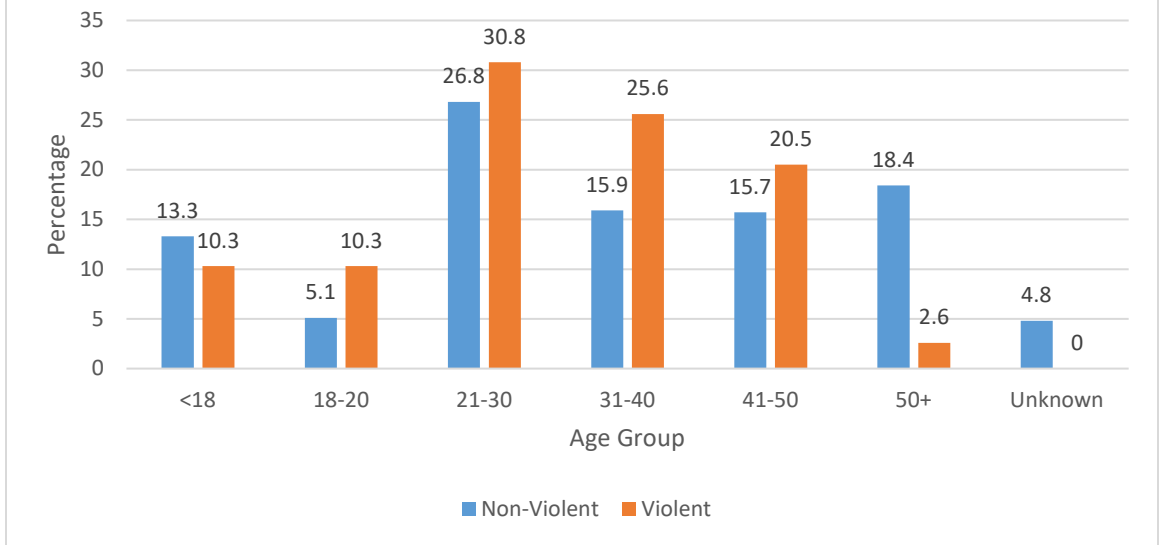


Figure 72. Comparison of Case Violence by Age Group at Missing Person Event, Males (N = 453)

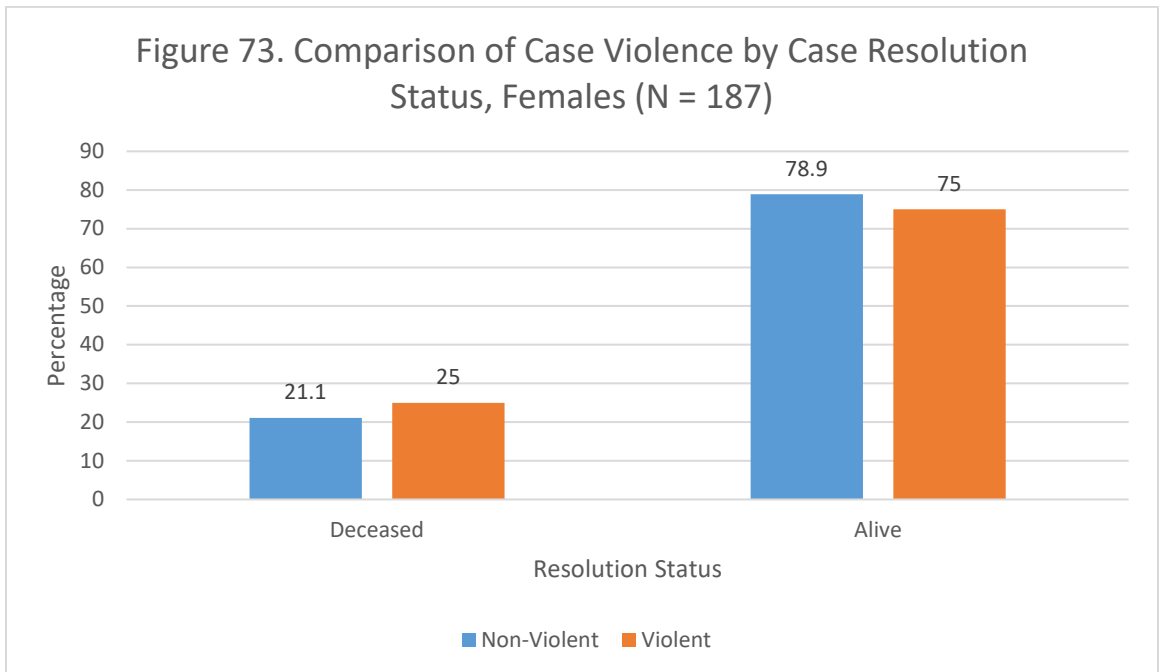


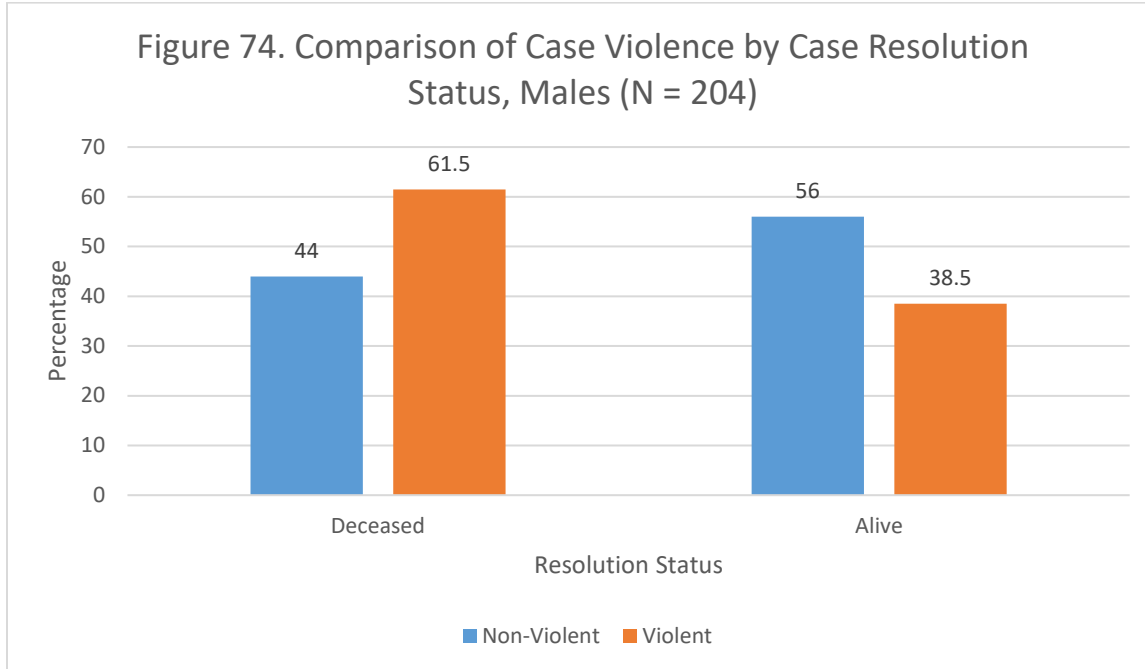
Resolved Cases (N = 391)

Case Resolution Status (N = 391)*

Among females, 21.1% of non-violent cases and 25.0% of violent cases had a resolution where the individual was found deceased (Figure 73). Among males, 44.0% (n = 84) of non-violent cases and 61.5% (n = 8) of violent cases had a resolution where the missing person was found deceased (Figure 74).

Figure 73. Comparison of Case Violence by Case Resolution Status, Females (N = 187)





NamUs Assisted (N = 390)*

Among females, 8.6% (n = 15) of non-violent cases and 8.3% (n = 1) of violent cases indicated NamUs assistance (Figure 75). For males, 16.3% (n = 31) of non-violent cases and 38.5% (n = 5) of violent cases indicated NamUs assistance (Figure 76).

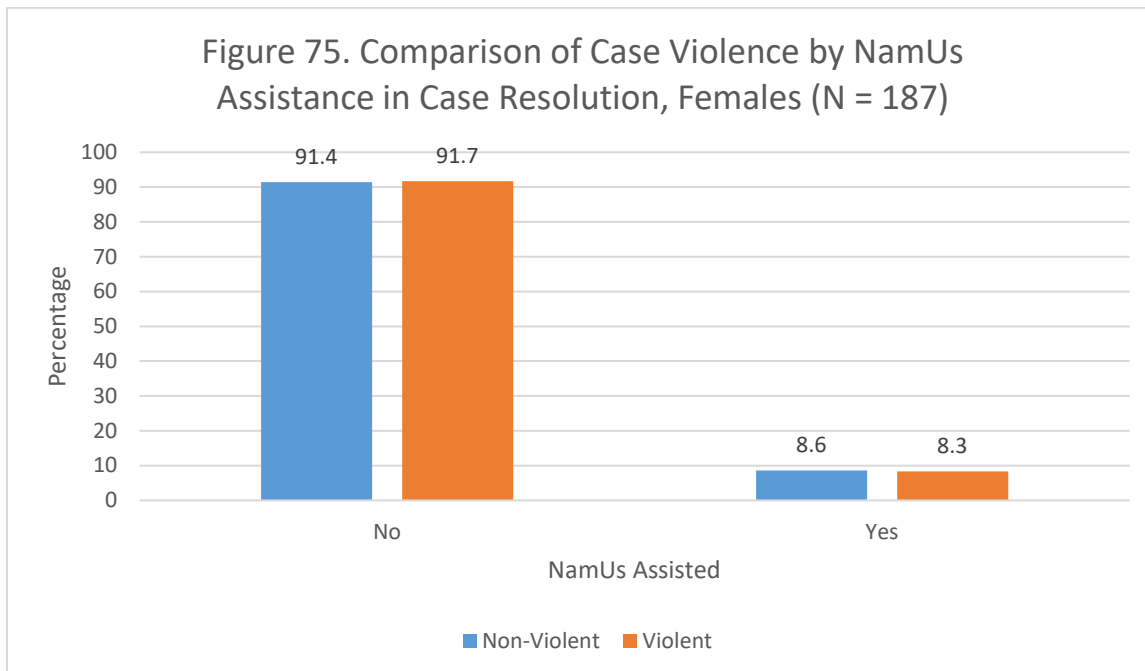
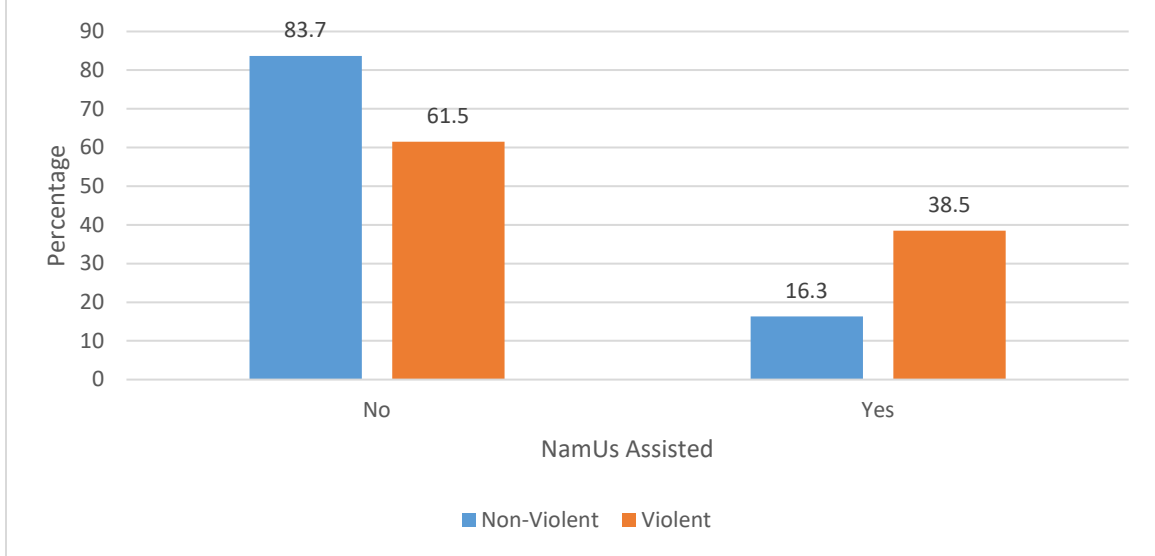


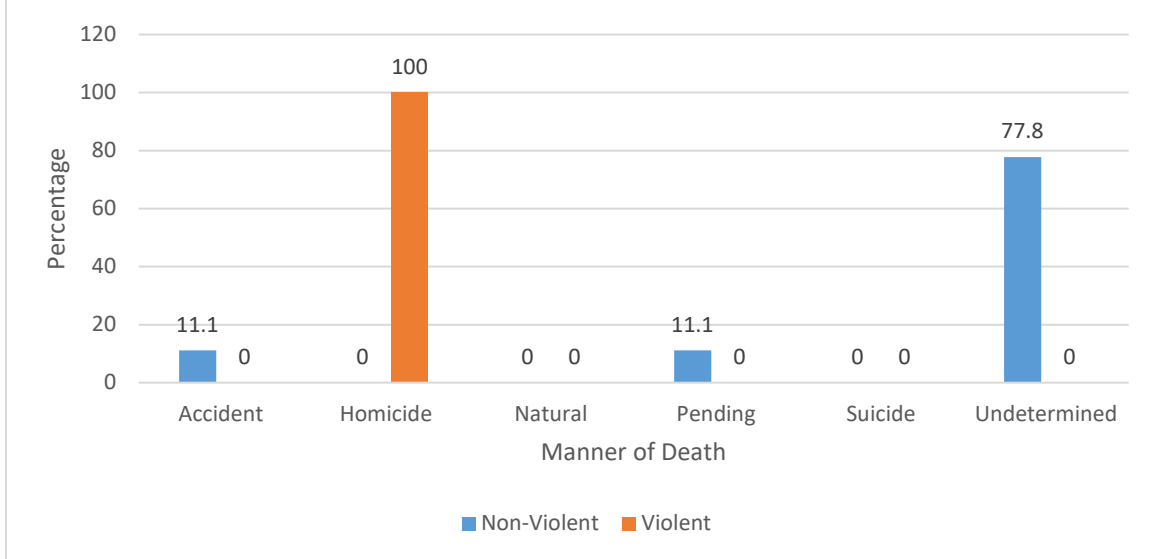
Figure 76. Comparison of Case Violence by NamUs Assistance in Case Resolution, Males (N = 203)

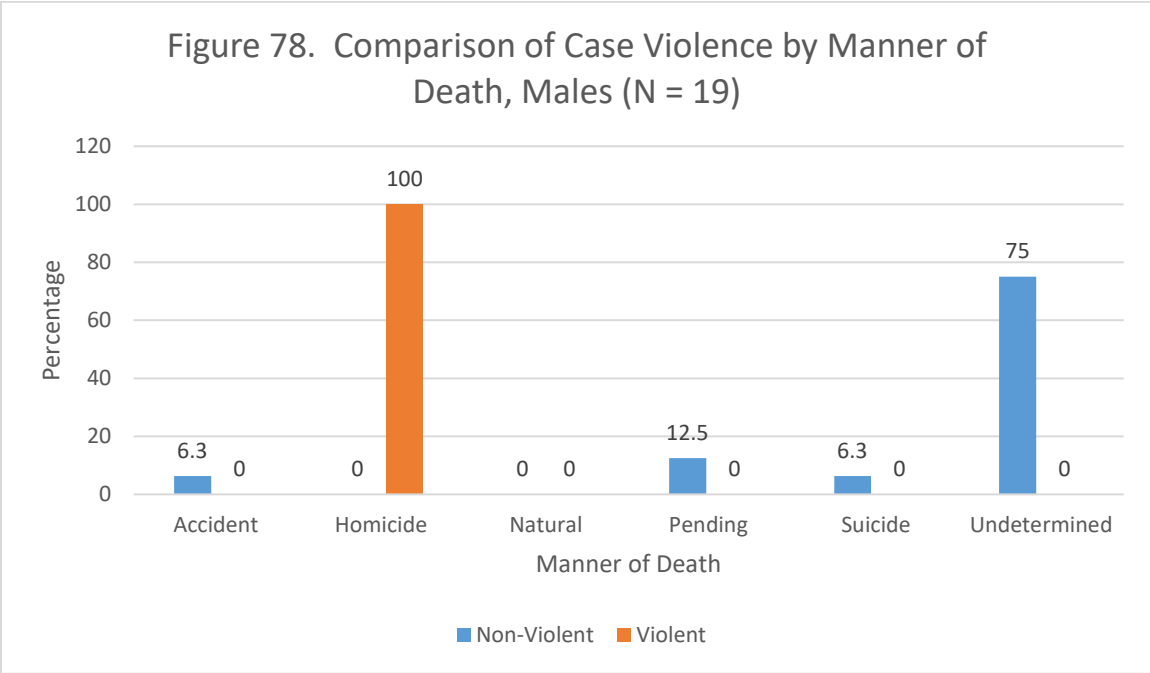


Manner of Death (N = 29)*

Among females classified as non-violent, the highest proportion of cases had undetermined as the MOD (77.8%, n = 7). Among female violent cases, 100.0% (n = 1) of cases were ruled as homicides (Figure 77). Among male non-violent cases, the highest proportion for manner of death was undetermined (75.0%, n = 12) for non-violent cases. For male violent cases, the most represented MOD was homicide (100.0%, n = 3) (Figure 78).

Figure 77. Comparison of Case Violence Across Manners of Death, Females (N = 10)

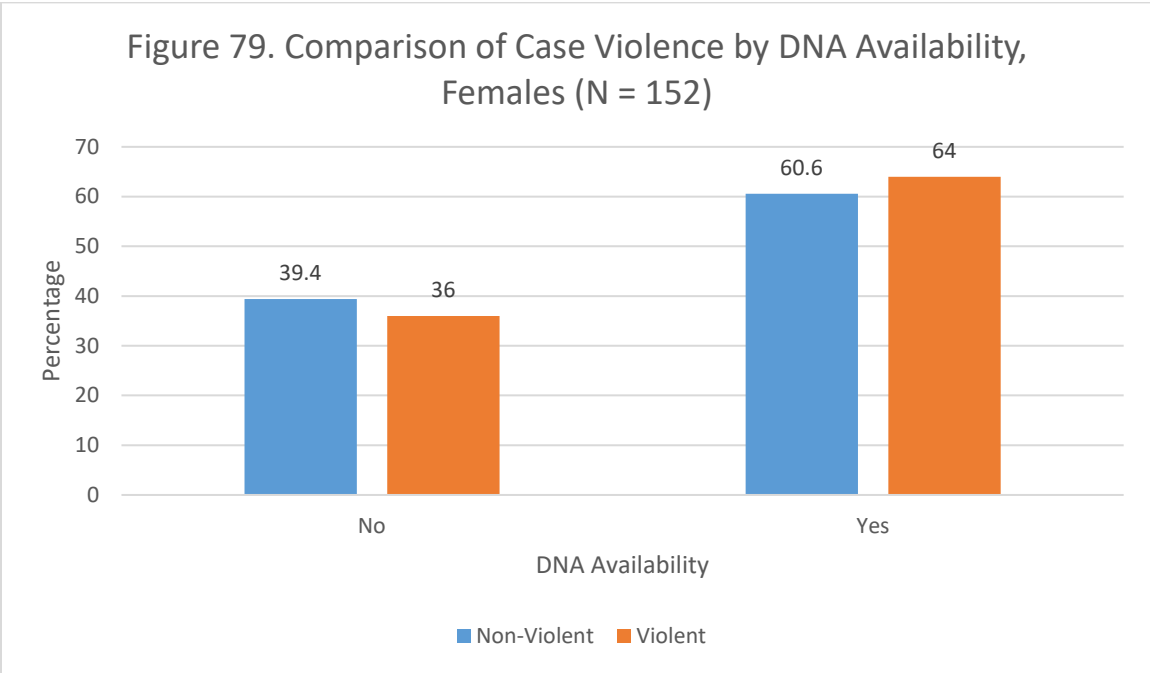


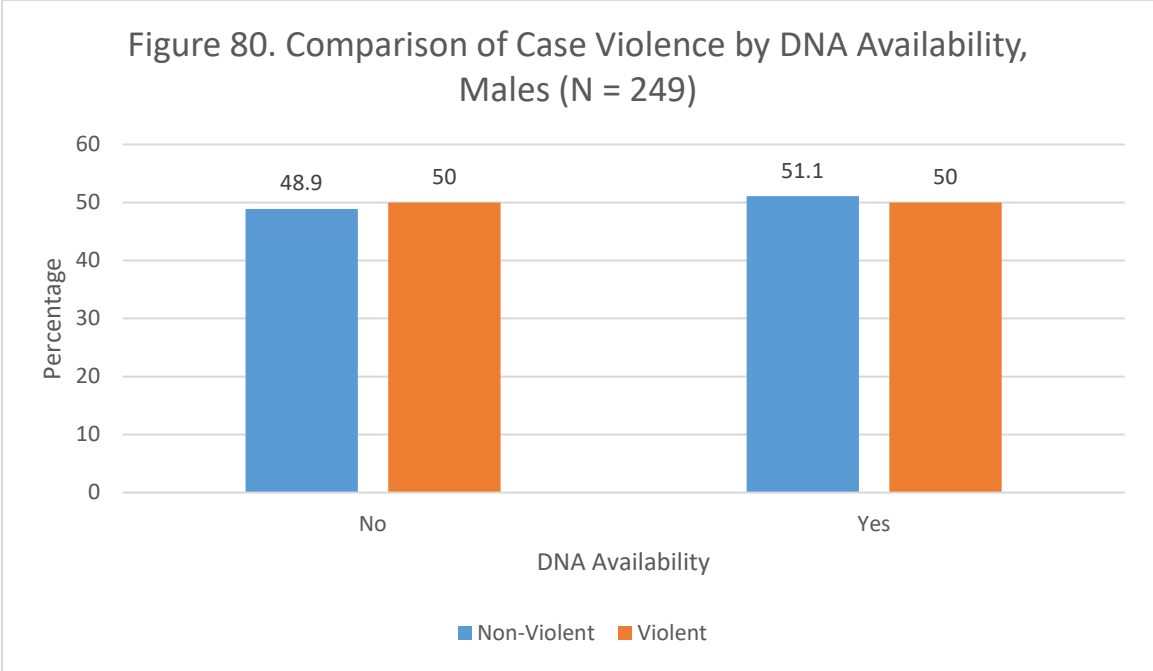


Unresolved Cases (N = 401)

DNA Availability (N = 401)*

Among female non-violent cases, 60.6% (n = 77) had DNA available while 64.0% (n = 16) of violent cases also had DNA available (Figure 79). Among males, 51.1% (n = 114) of non-violent cases and 50.0% (n = 13) of violent cases had DNA available for analysis (Figure 80).





Results: UP

Results for AI/AN UP case are not shown due to low cell counts. In most of the results, there is at least one cell with a cell count of less than 10. Given these low numbers, the AI/AN results were not included.

Recommendations

No recommendations are made for the AI/AN special section due to low numbers in the analyses. It was determined by the project team that providing recommendations for changes to the entirety of a national system based on results that often involved tables with cell counts less than 10 would not be appropriate. Once more cases have been entered into the database and the analysis can be re-run with higher case numbers, recommendations may be possible and appropriate.

Conclusion

Overall, results for MP AI/AN cases are presented. Results for AI/AN UP cases were not included due to low cell counts. Low cell counts also were observed with the MP cases. As such, no recommendations were made for systems improvement and expansion, as it was felt that it would be inappropriate to make broad generalizations for systems change based on several variables that had low cell count. However, the results for MP cases do provide preliminary insights into trends, which can be more fully described once more case data are captured within the system for MP and UP AI/AN cases.

Limitations

This tribal special section was limited in various ways. As with the general population analyses, the analysis presented in this section made use of operational data not designed specifically for the purposes of systems improvement or expansion inquiry. As such, the variables, including their codings, may not be the most appropriate data for the purposes of this report.

Second, the data fields used suffered from low numbers of cases with valid information, which limits the conclusions that can be drawn from any analysis involving these variables. As such, no recommendations for systems improvement and expansion are provided. Once more cases populate the NamUs database, it may be possible to make more general conclusions for systems improvement and expansion.

Finally, the cases used in these analyses represent self-identified AI/AN persons and not necessarily AI/AN persons enrolled or affiliated with a state- or federally-recognized tribe. As such, no inferences should be made regarding criminal jurisdiction.

V. General Conclusion

The work presented in this report led to noteworthy observations and conclusions that were useful in the development of potential hypotheses and recommendations for system improvement and expansion. Through this work, thoughtful and needed improvements or expansions to the system can be made informed by the relevant data. In all, this project provided valuable insights that may enable NamUs to better serve its constituent users, to more efficiently and effectively work to resolve cases, and, more broadly, to address the issues of missing and unidentified persons through prevention and intervention facilitated by better and more complete data collection, biometric sample collection, and service provision to persons impacted by such cases.

Summary of Recommendations

Overall, the work completed in this report resulted in several recommendations for NamUs improvement and expansion. In all, both the MP and UP sections resulted in five key findings with nine associated recommendations for a total of 18 recommendations for potential consideration for systems improvement and expansion. The implementation of these recommendations will depend on technological capabilities and human resources training to ensure that as many of the recommendations as possible are put in place. Some recommendations may require more data before considering possible implementation. Additionally, tracking and evaluation elements should be put in place to assess the impact of these recommendations and changes that occur that may have on case investigation, tracking, and resolution.

General Limitations

This project has several limitations. First, it uses data from a source that is not intended for the project purposes. The NamUs database is first and foremost an operational database used for the management, tracking, and resolution of cases. As such, data fields and responses are not optimized for the systems improvement purposes.

Second, the NamUs system relies on crowd sourced data; that is, NamUs relies on outside parties to enter data into the database and does not itself as a program actively solicit cases. Further, participation in the system is largely voluntary, as is the provision of data for many of the data fields. Thus, there is no single, comprehensive, or standardized data entry across all users. Without these elements, cases as data points may vary substantially in terms of data provided and the completeness of that data.

Third, NamUs launched NamUs 2.0 in May 2018. Since that date, continual updates have been made to NamUs. These dynamic updates may change the types of data that are collected, how it is collected, and how they are reported. For example, the tribal data fields were added in December 2018, and therefore, they do not represent complete data for all tribal cases that occurred before that date.

Next Steps

Project Phase 2

Following completion of case classification, a second phase of the current project may be considered to ascertain final case disposition for cases classified as violent. In this second phase, follow-up surveys may be conducted both electronically through e-mail and over the phone through a call to the investigating agency. This survey may include questions about the

perpetrator, victim-perpetrator relationship, and any criminal justice involvement due to the case, such as arrest or prosecution. The results of this phase may allow for recommendations to improve system data collection, as well as provide information on the characteristics of perpetrators and violent cases within NamUs that may be used to enhance case resolution, for example, by helping to build profiles of both victims and offenders that may assist in identifying potentially violent cases and in apprehending any associated violent offenders.

VI. References

- Bonny, E., Almond, L., & Woolnough, P. (2016). Adult missing persons: Can an investigative framework be generated using behavioral themes? *Journal of Investigative Psychology and Offender Profiling*, 13, 296-312.
- Breiding, M.J. (2014). Prevalence and characteristics of sexual violence, stalking, and intimate partner violence victimization—National Intimate Partner and Sexual Violence Survey, United States, 2011. *Morbidity and Mortality Weekly Report*, 63(8), 1-18.
- Breiding, M.J., Chen, J., & Black, M.C., (2014). *Intimate Partner Violence in the United States – 2010*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Accessed 26 Oct 2020 at https://www.cdc.gov/violenceprevention/pdf/cdc_nisvs_ipv_report_2013_v17_single_a.pdf.
- Crime Data Explorer. (2019). *All Violent Crime Offender vs. Victim demographics*. Accessed 17 July 2019 at <https://crime-data-explorer.fr.cloud.gov/explorer/national/united-states/crime/2007/2017>.
- Congress.gov. (2020, Oct. 10). *Savanna’s Act*. Washington, D.C.: 116th U.S. Congress. Accessed 5 Nov 2020 at <https://www.congress.gov/116/plaws/publ165/PLAW-116publ165.pdf>.
- Federal Bureau of Investigations. (2020). *2019 NCIC Missing and Unidentified Persons Statistics*. Washington, DC: U.S. Department of Justice, Federal Bureau of Investigations. Accessed 25 Oct 2020 at <https://www.fbi.gov/file-repository/2019-ncic-missing-person-and-unidentified-person-statistics.pdf/view>.
- Henderson, M., Henderson, P., & Kiernan, C. (2000). Missing persons: Incidences, issues and impacts. *Australian Institute of Criminology: Trends and Issues in Crime and Criminal Justice*, 144, 1-6
- Hickman, M.J., Hughes, K.A., Strom, K.J., & Roper-Miller, J.D. (2007). *Medical Examiners and Coroners’ Offices, 2004*. Washington DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Accessed 17 Jul 2020 at <https://www.bjs.gov/content/pub/pdf/meco04.pdf>.
- Hughes, K.A., (2007). *Unidentified Human Remains in the United States, 1980-2004*. Washington DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Accessed 17 Jul 2020 at <https://www.bjs.gov/content/pub/pdf/uhrus04.pdf>.
- National Center for Injury Prevention and Control. (2020). *Leading Causes of Death Reports, 1981 - 2018*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for

- Disease Control and Prevention, National Center for Injury Control and Prevention. Accessed 26 Oct 2020 at <https://webappa.cdc.gov/sasweb/ncipc/leadcause.html>.
- Operation Lady Justice. (n.d.). *The Presidential Task Force on Missing and Murdered American Indians and Alaska Natives*. Accessed 5 Nov 2020 at <https://operationladyjustice.usdoj.gov/>.
- Puzzanchera, C., Chamberlin, G., and Kang, W. (2020). "Easy Access to the FBI's Supplementary Homicide Reports: 1980-2016." Accessed 5 Nov 2020 at <https://www.ojjdp.gov/ojstatbb/ezashr/>
- Ritter, N. (2007). Missing persons and unidentified remains: The nation's silent mass disaster. *NIJ Journal*, 256, 2-7.
- Rosay, A.B. (2016). *Violence Against American Indian and Alaska Native Women and Men: 2010 findings from the National Intimate Partner and Sexual Violence Survey*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Accessed 4 Aug 2020 at <https://www.ncjrs.gov/pdffiles1/nij/249736.pdf>.
- Shweta, Bajpai, R.C., & Chaturvedi, H.K. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, 41(3), 21-27.
- Smith, S.G., Zhang, X., Basile, K.C., Merrick, M.T., Wang, J., Kresnow, M., & Chen, J. (2018). *The National Intimate and Sexual Violence Survey (NISVS): 2015 Data Brief – Updated Release*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Control and Prevention. Accessed 26 Oct 2020 at <https://www.cdc.gov/violenceprevention/pdf/2015data-brief508.pdf>.
- Sovereign Bodies Institute. (n.d.). *About the data*. Accessed 5 Nov 2020 at <https://www.sovereign-bodies.org/mmiw-database>.
- Sovereign Bodies Institute. (2020). *To' Kee Skuy' Soo Ney-Wo-Chek': I will see you again in a good way. Progress Report*. Accessed 5 Nov 2020 at https://2a840442-f49a-45b0-b1a1-7531a7cd3d30.filesusr.com/ugd/6b33f7_c7031acf738f4f05a0bd46bf96486e58.pdf.
- Straus, M.A., Hamby, S.L., Boney-McCoy, S., & Sugarman, D.B. (1996). The revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues*, 17(3), 283-316.
- Urban Indian Health Institute. (2018). *Missing and murdered indigenous women and girls*. Seattle, WA: Seattle Indian Health Board, Urban Indian Health Institute. Accessed 5 Nov 2020 at <file:///C:/Users/sph0085/Downloads/UIHI-Missing-and-Murdered-Indigenous-Women-and-Girls-Report-20191009.pdf>.

Violence Policy Center. (2016). *When men murder women: An analysis of 2014 homicide data*. Accessed 23 Oct 2018 at <http://www.vpc.org/studies/wmmw2016.pdf>.

World Health Organization. (2016 Nov). *Violence against women: Intimate partner and sexual violence against women*. Accessed 23 Oct 2018 at <http://www.who.int/mediacentre/factsheets/fs239/en/>

Appendix A

Technical Note

Purpose

The purpose of this technical note is to provide more nuanced and in-depth information into the methods underlying the results and recommendations of the ECC project. Information detailed within is a more comprehensive accounting for the technical steps taken as part of this report.

Case Classification

MP Case Classification

Run 1

Data delivered for Run 1 included 34,866 cases. Following delivery of data for Run 1, data were cleaned by removing any cases that did not have any information in either of the circumstance fields, that indicated some form of “unknown,” or that contained only numbers or symbols. This cleaning resulted in the removal of 537 cases leaving a total case count of 34,329 unique cases. After cleaning, data were imported into WordStat for analysis using the previously-built classification dictionary described in the respective MP methods section. This classification resulted in 5,054 cases being coded as violent.

Percent Agreement. Next, data were imported into Stata in order to draw a random sample of 10% of cases (N=3,433) for second coding by a human coder. Second coding resulted in 87% agreement between WordStat and the human coder. Overall, of the 449 disagreements between WordStat and the human coder, 325 cases (72.4%) were instances in which WordStat coded the case as violent, but the human coder did not. Alternatively, 124 cases (27.6%) of cases were instances in which WordStat coded the case as non-violent, but the human coder coded the case as violent.

Disagreements and Dictionary Changes. After second coding, a disagreements check was performed in order to evaluate the relevance and functionality of keywords that led to discrepant coding between WordStat and the human coder. Following the disagreements check, appropriate amendments to the dictionary were made, including the addition of new keywords or inflected forms including, but not limited to: argue, abduct, homicide, exploitation, disagreement, killer, kidnapping, and danger. Some words were re-classified to more appropriate categories of violence, including foul play, riot, and brawl from general violence to physical violence. Further, the term “trafficking” was specified to be “gender trafficking” as it was felt that not all forms of trafficking could be adequately captured within the current dictionary and that a broader conversation on trafficking as an issue of interest to stakeholders was warranted before finalizing any further trafficking keywords. Finally, an exclusion list for keywords and phrases that were consistently resulting in false positives was created including but not limited to words and phrases such as: “substance abuse,” “follow-up,” “Air Force,” “self-harm,” and “bipolar disorder” among others. A full listing of excluded words and phrases is shown in Table 1.

Table 1. Exclusion List

Keyword	Excluded Phrase
Abuse	Abuse alcohol Abuse drugs Abuse substances Abuses alcohol

	Abuses drugs
	Abuses substances
	Alcohol abuse
	Alcohol abuser
	Substance abuse
	Substance abuser
	Drug abuse
	Drug abuser
Harm	Self-harm
Follow	Follow up
	Follow-up
	Followed-up
	Followed up
Disorder	Removed from dictionary

Data Discrepancies. Through the disagreement checking process, two data inconsistencies were discovered. First, 18 cases were identified in which some of the lines related to the cases were not present. There were likely several more cases with this issue. This issue made disagreement checking more difficult, as often the score provided by the computer was not substantiated by the single line of text but was validated upon checking the case within the original data. While the exact cause of this issue is unknown, it is likely that it occurred during the exportation process from either WordStat or Stata after the random sample of cases was selected. It is possible that not all lines were selected and exported as part of the random sample. Alternatively, it is possible that the process through which multiple-line cases were collapsed and summed to check for agreement with the human coded data was faulty. This issue was not discovered initially, as there was no reason to believe cases would not be exported in whole by CaseID. Although there were data inconsistencies, it is not believed that the inter-coder agreement was affected, as the exact same information was coded by both the computer and the human. In the second run of the data, particular attention was paid to any import and export processes to ensure that no cases were lost in part or in full. Additionally, it was attempted to use Stata as little as possible and to work within the spreadsheets manually whenever possible.

Run Termination. Following these steps, however, we eventually terminated this particular run due to missing Circumstances of Resolution data for text analysis. These data had not been requested by the project team initially but were later determined by to be important for case classification. As such, no tertiary coding was conducted as part of this run.

Run 2

Data delivered for the second classification included again 34,866 cases and also included Circumstances of Resolution in addition to other circumstance fields. These data were merged with general Circumstances and Agency Notes fields to create the full set of data fields that would be classified. Data were cleaned using the same process as in Run 1. This cleaning resulted in the removal of 356 cases in their totality and one case with a missing Case ID resulting in a final sample of 34,509 unique cases for classification. Cleaned data were then imported into WordStat for classification. Several classifications were tested, including with and without an exclusion list created based on lessons learned from Run 1 of the data.

Misspellings Check. In Run 2, we also added a misspellings check to the procedures. This check was completed using WordStat’s built-in functionality. The misspellings check was conducted using the lowest level of confidence (i.e., moderate) of a potential misspelling for words in the classification dictionary to ensure the broadest set of potential misspelled words was captured and assessed. Each potential misspelling was checked using a “keyword-in-context” feature to ensure the misspelling truly represented a misspelled word, a proper noun, or some otherwise correct spelling for a different word. Overall, 134 potentially misspelled words were identified by WordStat. Upon checking the keyword-in-context, it was determined that 40 instances represented valid misspellings and were added within the classification dictionary. The other instances were generally proper nouns, names, or parts of internet addresses and were not added to the classification dictionary.

Percent Agreement. The final run of the data included use of the exclusions list previously created, as well as corrected misspelled words. Overall, this classification resulted in 7,059 cases being coded as violent by WordStat. This number increased by nearly 2,000 cases from Run 1, likely due to the addition of keywords (e.g., beat, argue, homicide, prostitution), as well as the inclusion of Circumstances of Resolution as a text field for analysis. Following the final Run 2, data were imported into Stata, and a random sample of 10% (N=3,451) was drawn for second coding. Percent agreement was calculated and was found to be 89%. Of the 383 disagreements, 343 (89.6%) were found to be instances in which WordStat classified a case a violent, but the human coder did not. Alternatively, 40 (10.4%) cases were instances in which the human coder classified the case as violent, but WordStat did not.

Disagreements and Dictionary Changes. Following the calculation of percent agreement, a disagreements check was performed to understand discrepancies in coding between WordStat and the human coder. Through this process, no additional keywords were uncovered to be added to the dictionary. However, through review of the dictionary by co-investigator Spamer, it was determined that there were additional words and their inflected forms that needed to be added to the dictionary. Additional keywords added included: pimp, prostituting, sexual assault, suspect, confession, offender, gunshot, strangle, asphyxiate, ligature, knifepoint, bury, smother, manslaughter, guilty, and sentence. As such, it was determined that a third run of the data would be necessary.

Tertiary Coding. No tertiary coding was conducted as part of this run of the data.

Run 3

Baseline data were imported from Run 2, as the cleaning procedure was already complete for the full dataset that included Circumstances of Resolution. New keywords were received from co-investigator Spamer following Run 2 were integrated into the dictionary. Overall, 7,659 (22.2%) of cases had at least one keyword indicating violence. This number increased by 600 cases from Run 2, likely due to the addition of keywords by co-investigator Spamer.

Misspellings Check. Then, in WordStat, a misspellings check was conducted in the same manner as before, and all identified misspellings that represented truly misspelled words were categorized into the WordStat dictionary. Overall, 108 potential misspellings were identified by WordStat. Of those potential misspellings, nine were valid and were added to the classification dictionary. The non-valid misspellings represented proper nouns, names, and locations, among other categories.

Percent Agreement. Following classification in WordStat, a random sample of 10% (N=3,451) for second coding was drawn in Stata. Following WordStat coding, percent

agreement between WordStat and the human coder was calculated and found to be 88%. Overall, WordStat and the human coder disagreed on 401 cases. Of these disagreements, 389 (97%) represented cases where WordStat classified the case as violent, but the human coder did not. Alternatively, 12 (3.0%) cases were classified as violent by the human coder, but as non-violent by WordStat.

Tertiary Coding. Given the high level of agreement between WordStat and the human coder, tertiary coding was completed by a second human coder (co-investigator Spamer) using a subsample of 20% of the 10% of cases originally sampled for second coding (N=690). The agreement between the two human coders was 93%. Overall, the human coders disagreed on 48 cases. Six cases were coded violent by Spamer, but as non-violent by Hafner, accounting for 12.5% of all disagreements. Alternatively, 42 cases were coded as violent by the Hafner, but as non-violent Spamer, accounting for 87.5% of disagreements.

Final Violence Coding

Following tertiary coding, given the high levels of agreement between WordStat and the human coder, as well as between the two human coders, a final coding of all cases classified as violent by WordStat was performed. In doing so, only cases classified as violent by WordStat were examined to ensure they truly represented violence and that the classification as violent was not spurious based on keywords being used in a different context (e.g., DNA “hit”). Overall, there were 7,659 cases from Run 3 coded as violent by WordStat, but only 3,493 were retained as actually indicating suspected or confirmed violence by the human coder. Later, 86 cases with a manner of death of homicide that were not classified in the case classification as violent were added as violent cases resulting in the final case count of 3,579 violent cases.

UP Case Classification

Run 1

Run 1 consisted only of cases in which the manner of death was undetermined or for which no manner of death was provided. Data delivered for this first run included 16,857 cases overall with 6,6617 undetermined cases and 2,581 cases where the manner of death was unknown, totaling 9,198 cases for classification. Following delivery of data, the data were cleaned by removing any cases that did not have any information in either of the circumstance fields, that indicated some form of “unknown,” or that contained only numbers or symbols. This cleaning resulted in the removal of 103 cases leaving a total case count of 9,095 unique cases. After this initial cleaning, data were imported into WordStat for analysis using the previously-built classification dictionary described above.

Misspellings. As with MP cases, we used WordStat’s built-in misspelling functionality to check for potentially misspelled yet valid keywords. Overall, 11 potential misspellings were identified using the lowest level of confidence within the WordStat functionality. In the end, nine of these misspellings were found to be valid and were included in the classification dictionary.

Percent Agreement. Next, data were imported into Stata in order to draw a random sample of 10% of cases for second coding by a human coder. During the second coding, a case that should have been cleaned out was discovered and was removed leaving an overall sample of 9,094 cases. However, after discovery of the invalid case described, the removal of it resulted in a second coding sample of 909 valid cases. Second coding resulted in 87.2% agreement between

WordStat and the human coder. Of the 116 total disagreement cases, 102 (87.9%) were cases in which WordStat classified the case as violent, but the human coder did not. Fourteen cases (12.1%) were instances in which the human coder classified the case as violent, but WordStat did not. Overall, following this misspelling check and additional cleaning, the classification resulted in 1,244 cases being coded as violent.

Disagreements Check. A disagreements check was performed in order to evaluate the relevance and functionality of keywords that led to discrepant coding between WordStat and the human coder. Following the disagreements check, it was determined that all words were relevant to the analysis and that even though some words (e.g, trauma) might result in several false positives that such particular keywords were too central to the violent case classification to be removed.

Tertiary Coding. After second coding, tertiary coding was conducted by a second human coder. For this coding, a subsample of 20% of the 10% used for second coding was drawn (n = 182) and was delivered to the tertiary coder. However, during the tertiary coding process, truncated circumstances were discovered by the second human coder.

Run Termination. Given the truncated circumstances uncovered, we eventually aborted this first run due to the truncated circumstance data. As such, the data used for this run were no longer considered valid and were discarded.

Run 2

Data delivered for the second classification included the same number of cases as Run 1 (N = 9,198). Initial cleaning resulted in the removal of 113 cases and a final sample of 9,085 unique cases for classification. Cleaned data were then imported into WordStat for classification. However, during the second coding process described below, two additional cases were found that were cleaned out from the data resulting in 115 removals and a unique sample of 9,083 cases.

Misspellings Check. We conducted a misspellings check that was completed using WordStat's built-in functionality, and each potential misspelling was checked using a "keyword-in-context" feature to ensure the misspelling truly represented a misspelled word, a proper noun, or some other otherwise correct spelling for a different word. Overall, 18 potential misspelled words were found, and 11 words were found to be valid misspellings of a keyword. These 11 misspellings were included in the classification dictionary.

Percent Agreement. Following the final Run 2, which included corrected misspelled words, this classification run resulted in 1,256 cases being coded as violent by WordStat. Data were then imported into Stata, and a random sample of 10% of all cases (n = 908) was drawn for second coding. Initial percent agreement was calculated and was found to be 87.2%.

Disagreements and Dictionary Changes. Following the calculation of percent agreement, a disagreements check was performed to understand discrepancies in coding between WordStat and the human coder. Through this process, a discrepancy was discussed among coders for the word "fetus." It was determined that just because that decedent was a fetus, infant, baby, or similar, that it could not be assumed that nefarious actions had led to its death. Thus, changes were made to classifications of 13 cases in which fetus or a similar keyword caused a violent classification to a non-violent classification. This change resulted in a new percent agreement between WordStat and the human coder of 88.2%. Of the 107 disagreements, 94 (87.9%) were cases in which the WordStat classified a case as violent but the human coder did not. These types of disagreements occur when a keyword is located in the text being analyzed,

but in context, the keyword is not indicating violence in the intended way. The other 13 (12.1%) disagreements were cases in which the human coder coded the case as violent, but there was no keyword found by WordStat in the text.

No changes to the classification dictionary were made based off of the disagreements from Run 2.

Tertiary Coding. For tertiary coding, a subsample of 20% of the 10% sample drawn for second coding was drawn at random in Stata (n = 182). Tertiary coding resulted in a percent agreement of 97.8% between human coders.

Final Violence Coding. Following secondary and tertiary coding, a manual coding of all cases indicated as violent by WordStat was conducted to verify the violent nature of the case. In all, 1,256 cases were re-evaluated. Resulting from this process were 96 valid cases of case violence that were retained in later analyses as violent cases, having met the threshold to be deemed as truly violent cases. This threshold typically involved explicit statements that the case was possibly, probably, or definitely a homicide. Inferences of violence (e.g., shallow grave) did not meet the threshold for violent classification.

Run 3

Following discussion with co-investigator Spamer, it was determined that “pending” manner of death cases should also be classified as violent or non-violent. As such, full data representing 16,857 cases, of which 1,064 were pending cases, were imported for classification. Data cleaning resulted in one case being removed for a new total case count of 1,063 pending cases.

Misspellings Check. Then, in WordStat, a misspellings check was conducted in the same manner as before, and all identified misspellings that represented truly misspelled words were categorized into the WordStat dictionary. Overall, only two possible misspellings were identified, and neither was found to be valid.

Percent Agreement. Overall, WordStat classification resulted in 163 cases in which a violent keyword was present. Following classification in WordStat, a random sample of 10% (N = 106) for second coding was drawn in Stata. Following WordStat coding, percent agreement between WordStat and the human coder was calculated and found to be 82%.

Disagreements and Dictionary Changes. Overall, there were no consistently apparent keywords leading to misclassifications. “Trauma” and “shallow grave,” did appear a few times, as did “gunshot,” “burned,” and “foul play,” but all keywords were too central to the violent classification dictionary to result in any changes to the dictionary.

Tertiary Coding. Given the level of agreement between WordStat and the human coder, tertiary coding was completed by a second human coder (co-investigator Spamer) using a subsample of 20% of the 10% of cases originally sampled for second coding (n = 21). The agreement between the two human coders was 95.2%.

Final Violence Coding. Following tertiary coding, a final coding of all cases classified as violent by WordStat was performed. Only cases classified as violent by WordStat (n = 163) were examined to ensure they truly represented violence and that the classification as violent was not spurious based on keywords being used in a different context (e.g., DNA “hit”). Of the 163 cases coded as violent by WordStat, 14 cases were retained as actually indicating suspected or confirmed violence by the human coder. The threshold for retention typically involved explicit statements that the case was possibly, probably, or definitely a homicide. Assumptions of violence (e.g., shallow grave) did not meet the threshold for violent classification.

Final UP Violence Coding

Overall, there were 16,857 cases under consideration for the UP analysis. 6,574 cases were undetermined manner of death, 2,509 cases did not have a manner of death provided, and 1,064 cases had a pending manner of death at the time the data were pulled. From text mining classification, 110 total cases were found to indicate confirmed or suspected violence (96 cases from undetermined or no manner of death, 14 cases from pending manner of death).

Case Re-Classification. During the final coding process, it was noticed that the inflected forms for “abduct” and “suspicious” were not completely outlined. Missing were the inflected forms “abduction” and “suspicion.” Overall, 4 cases of “suspicion” and 47 cases of “abduction” were discovered. The classification of the associated cases was thus changed from non-violent to violent. In total, the original and re-classifications resulted in 2,655 cases being coded as violent either through a homicide determination or through the case classification process.

MP Case Analysis

Analytical Variables

Table 1 outlines the analytical variables used as part of case analysis, including their original coding as delivered by NamUs and the steps taken to code them in their final form.

Violence Variables			
Variable	Original Coding	Re-code	Notes
Finalviolence	Yes No	0 = No 1 = Yes	Copied “finalhumancode” Dichotomized from “finalhumancode” copy
GV (GENERAL_VIOLENCE)	0 1 2 3 4 5 6 7 8 9 10 11 12 16 18 39	0 = No 1 = Yes	Copied “General_Violence” to create “GV” Replaced “GV” = 0 if finalviolence=0 Dichotomized from “GV”
PV (PHYSICAL_VIOLENCE)	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 17 23	0 = No 1 = Yes	Copied “Physical_Violence” to create “PV” Replaced “PV” = 0 if finalviolence=0 Dichotomized from “PV”
SV (SEXUAL_VIOLENCE)	0 1 2 3	0 = No 1 = Yes	Copied “Sexual_Violence” to create “SV” Replaced “SV” = 0 if finalviolence=0

	4 6		Dichotomized from "SV"
ST (STALKING)	0 1 2 3 4 5 6 8 9 10	0 = No 1 = Yes	Copied "Stalking" to create "ST" Replaced "ST" = 0 if finalviolence=0 Dichotomized from "ST"
CC (COERCIVE_CONTROL)	0 1 2 3 4 5 6 8 10 23	0 = No 1 = Yes	Copied "Coercive_Control" to create "CC" Replaced "CC" = 0 if finalviolence=0 Dichotomized from "CC"
PA (PSYCHOLOGICAL_AGGRESSION)	0 1 2 3 4 5 6	0 = No 1 = Yes	Copied "Psychological_Aggression" to create "PA" Replaced "PA" = 0 if finalviolence=0 Dichotomized from "PA"
All Case Characteristics			
Variable	Original Coding	Re-code	Notes
Age_of_Case_groups2	<90 days 3-12 mo. 1-2 years 2-5 years 5-10 years 10-20 years 20+ years Unknown	0 = <90 days 1 = 3-12 mo. 2 = 1-2 years 3 = 2-5 years 4 = 5-10 years 5 = 10-20 years 6 = 20+ years 7 = Unknown	Encode of "Age_of_Case_groups" Re-labeled due to re-ordering of categories
StateName2	Alabama Alaska Arizona Arkansas California Colorado Connecticut Delaware DC Florida Georgia Guam Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana Maine Maryland Massachusetts Michigan Minnesota Mississippi Missouri Montana Nebraska Nevada	1 = Alabama 2 = Alaska 3 = Arizona 4 = Arkansas 5 = California 6 = Colorado 7 = Connecticut 8 = Delaware 9 = DC 10 = Florida 11 = Georgia 12 = Guam 13 = Hawaii 14 = Idaho 15 = Illinois 16 = Indiana 17 = Iowa 18 = Kansas 19 = Kentucky 20 = Louisiana 21 = Maine 22 = Maryland 23 = Massachusetts 24 = Michigan 25 = Minnesota 26 = Mississippi 27 = Missouri 28 = Montana 29 = Nebraska 30 = Nevada	Encode of "StateMissing" Re-labeled due to re-ordering of categories

	New Hampshire New Jersey New Mexico New York North Carolina North Dakota Northern Mariana Islands Ohio Oklahoma Oregon Pennsylvania Puerto Rico Rhode Island South Carolina South Dakota Tennessee Texas Utah Vermont Virgin Islands Virginia Washington West Virginia Wisconsin Wyoming	31 = New Hampshire 32 = New Jersey 33 = New Mexico 34 = New York 35 = North Carolina 36 = North Dakota 37 = Northern Mariana Islands 38 = Ohio 39 = Oklahoma 40 = Oregon 41 = Pennsylvania 42 = Puerto Rico 43 = Rhode Island 44 = South Carolina 45 = South Dakota 46 = Tennessee 47 = Texas 48 = Utah 49 = Vermont 50 = Virgin Islands 51 = Virginia 52 = Washington 53 = West Virginia 54 = Wisconsin 55 = Wyoming	
FoulPlay2	No Possibly Uncertain Yes	0 = No 1 = Possibly 2 = Uncertain 3 = Yes	Encode of "FoulPlay"
SexName2	Female Male Other	0 = Female 1 = Male 2 = Other	Encode of "SexName" Re-labeled due to re-ordering of categories
RaceEthnicity2	[Multiple Categories]	0 = White / Caucasian 1 = American Indian / Alaska Native 2 = Asian 3 = Black / African American 4 = Hawaiian / Pacific Islander 5 = Hispanic / Latino 6 = Other 7 = Uncertain 8 = 2+ Races	Encode of "RaceEthnicity" Re-labeled due to re-ordering of categories Original coding of individuals races checked through "codebook" of individual categories: 1=AI/AN 8=Asian 17=Black/AA 28=Hawaiian/PI 31=Hispanic/Latino 35=Other 38=Uncertain 40=White/Caucasian
Age_Missing_Group2	18-20 21-30 31-40 41-50 50+ <18 Unknown	0 = <18 1 = 18-20 2 = 21-30 3 = 31-40 4 = 41-50 5 = 50+ 6 = Unknown	Encode of "Age_Missing_Group"
Resolved Case Characteristics			
Variable	Original Coding	Re-code	Notes
MannerOfDeath2	Accident Homicide Natural Pending Suicide Undetermined	1 = Accident 2 = Homicide 3 = Natural 4 = Pending 5 = Suicide 6 = Undetermined	Encode of "MannerOfDeath"
CaseResolutionStatus2	Alive Deceased	0 = Deceased 1 = Alive	Encode of "CaseResolutionStatus" Re-labeled due to re-ordering of categories
NamUsAssisted	0 1	0 = No 1 = Yes	No "encode" necessary as numeric already

			Chosen over NamUsAssisted_YN, because that variable put all missing data as Yes
Resolved Case Characteristics			
Variable	Original Coding	Re-code	Notes
DNA_Indicator2	N Y	0 = No 1 = Yes	Encode of "DNA_Indicator" Re-labeled due to re-ordering of categories

Missing Cases

During the analysis of cases, it was noticed that across several variables, some of which were mandatory fields, there was consistently four cases missing. After consulting the data scientist who provided the data, it was uncovered that there had been a mistake in assembling the dataset such that four specific cases were listed with shifted rows, which created the four case deficit. The data scientist provided corrected data for these cases, and these cases were then manually re-coded to reflect the true answers for the cases within the full dataset.

Case Re-Classifications

Initially, resolved cases were not classified as violent or non-violent based on their manner of death. However, after consultation with co-investigators, it was determined that cases with a listed manner of death of "homicide" should be forced into the analysis as violent cases even when circumstantial information mined did not indicate any foul play or violence. In doing so, 86 cases that were identified as having "homicide" as their manner of death, but which lacked circumstantial information indicating violence, were moved to be violent cases for the remainder of the classification.

UP Case Analysis

Analytical Variables

Table 1 outlines the analytical variables used as part of case analysis, including their original coding as delivered by NamUs and the steps taken to code them in their final form.

All Case Characteristics			
Variable	Original Coding	Re-code	Notes
finalviolence	Variable created by research team	0 = No 1 = Yes	Based on case classification, as well as medical examiner/coroner rulings of "homicide" manner of death
StateName2	[Multiple Values]	None	Encode of "StateName"
ConditionOfRemainsName2	Not Recognizable (8 categories) Recognizable	0 = Not recognizable 1 = recognizable face	Encode of "ConditionOfRemains" Re-labeled due to new categories
Age of Case Groups	0-1 Years 1-5 years 5-10 years 10-20 years 20+ years Unknown	1 = 0-1 Years 2 = 1-5 years 3 = 5-10 years 4 = 10-20 years 5 = 20+ years 6 = Unknown	Encode of "Age_of_Case_groups" No re-label – already in correct order
SexName2	Female Male Unsure	0 = Female 1 = Male 2 = Unsure	Encode of "SexName" Re-labeled due to change of categories
RaceEthnicity2	[Multiple Categories]	0 = White / Caucasian 1 = American Indian / Alaska Native 2 = Asian 3 = Black / African American	49 unique values Encode of "RaceEthnicity"

		4 = Hawaiian / Pacific Islander 5 = Hispanic / Latino 6 = Other 7 = Uncertain 8 = 2+ Races	
AgeGroup	Based on median age of case	0 = <18 1 = 18-20 2 = 21+	Used data only from cases where both minimum and maximum age were provided to calculate a median age
Resolved Case Characteristics			
Variable	Original Coding	Re-code	Notes
NamUsAssisted	0 1	0 = No 1 = Yes	Already numeric, no encode allowed
MannerOfDeathName3	Accident Homicide Natural Pending Suicide Undetermined	1 = Accident 2 = Homicide 3 = Natural 4 = Pending 5 = Suicide 6 = Undetermined	Encode of "MannerOfDeathName2"
MethodOfId_Anthropology	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_CircumstantialInforma	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_Dental	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_Fingerprints	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_MtDNA	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_NucDNA	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_Other	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_Radiograph	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
MethodOfId_VisualIdentification	0 1	0 = No 1 = Yes	Already numeric, no encode Added labels
Unresolved Case Characteristics			
Variable	Original Coding	Re-code	Notes
DNA_Indicator2	No Yes	0 = No 1 = Yes	Encode of "DNA_Indicator"

Appendix B

Final Classification Keywords

Final MP Classification Keywords

General Violence (46)

- Violence
- Aggression
- Attack
- Assault
- Threat
- Abuse
- Rampage
- Conflict
- Dispute
- Mistreat
- Maltreat
- Battle
- Skirmish
- Kerfuffle
- Ruckus
- Fracas
- Fury
- Hubbub
- Hostile
- Brouhaha
- Tirade
- Brutal
- Row
- Disorder
- Cruel
- Scuttle
- Discriminate
- Anger
- Torture
- Neglect
- Endanger
- Bully
- General Violence
- Fight
- Suspicious
- Confront
- Altercation
- Danger
- Offender
- Confess
- Bury
- Grave
- Guilty
- Sentence

- Gang
- Suspect

Psychological Aggression (17)

- Yell
- Shout
- Scream
- Insult
- Name Call
- Humiliate
- Made Fun Of
- Cyberbully
- Cursed At
- Psychological Aggression
- Verbal Fight
- Verbal Assault
- Intimidate
- Verbal Disagreement
- Argue
- Disagreement
- Verbal Altercation

Physical Violence (40)

- Hit
- Slap
- Push
- Shove
- Punch
- Kick
- Stab
- Shoot
- Bite
- Burn
- Pull
- Destroy
- Pull Hair
- Slam
- Choke
- Suffocate
- Weapon
- Twisted Arm
- Grab
- Knock Down
- Kill
- Murder

- Abduct
- Kidnap
- Physical Violence
- Physical Fight
- Physical Assault
- Rumble
- Harm
- Injure
- Homicide
- Beat
- Brawl
- Riot
- Foul Play
- Physical Confrontation
- Strangle
- Asphyxiate
- Smother
- Manslaughter

Coercive Control (10)

- Force
- Coerce
- Manipulate
- Isolated
- Coercive Control
- Control
- Sex Traffick
- Exploit
- Pimp
- Human Trafficking

Sexual Violence (5)

- Rape
- Penetrate
- Molest
- Sexual Violence
- Sexual Assault

Stalking (6)

- Stalk
- Follow
- Spy
- Watch
- Broke In
- Unwanted Communication

Final UP Classification Keywords

- Sexual Assault
- Rape
- Suspect
- Offender
- Confess
- Bury
- Shot
- Strangle
- Asphyxiate
- Ligature
- Knife
- Stab
- Grave
- Smother
- Manslaughter
- Guilty
- Sentence
- Burn
- Trauma
- Blunt Force
- Bludgeon
- Torture
- Suspicious
- Foul Play
- Dismember
- Choke
- Suffocate
- Murder
- Homicide
- Beat
- Attack
- Assault
- Hit
- Punch
- Fight
- Gun
- Kill