

Original Paper

# Automated Analysis of Domestic Violence Police Reports to Explore Abuse Types and Victim Injuries: Text Mining Study

George Karystianis<sup>1</sup>, BSc, MSc, PhD; Armita Adily<sup>1</sup>, BMedSc, MPH, PhD; Peter W Schofield<sup>2</sup>, BSc (Hons), BS, MSc, MD; David Greenberg<sup>3</sup>, MB, ChB, FFPsych (SA), FRCPC, FRANZCP, MMed; Louisa Jorm<sup>4</sup>, BVSc, MSc, PhD; Goran Nenadic<sup>5</sup>, BSc, MSc, PhD; Tony Butler<sup>1</sup>, BSc, MSc, PhD

<sup>1</sup>The Kirby Institute, Faculty of Medicine, The University of New South Wales, Sydney, Australia

<sup>2</sup>Neuropsychiatry Service, Hunter New England Health, Newcastle, Australia

<sup>3</sup>School of Psychiatry, The University of New South Wales, Sydney, Australia

<sup>4</sup>Centre for Big Data Research in Health, The University of New South Wales, Sydney, Australia

<sup>5</sup>School of Computer Science, The University of Manchester, Manchester, United Kingdom

**Corresponding Author:**

George Karystianis, BSc, MSc, PhD

The Kirby Institute

Faculty of Medicine

The University of New South Wales

Level 6, Wallace Wurth Building

High Street, Kensington NSW

Sydney, 2052

Australia

Phone: 61 (2) 9385 0900

Email: [gkarystianis@kirby.unsw.edu.au](mailto:gkarystianis@kirby.unsw.edu.au)

## Abstract

**Background:** The police attend numerous domestic violence events each year, recording details of these events as both structured (coded) data and unstructured free-text narratives. Abuse types (including physical, psychological, emotional, and financial) conducted by persons of interest (POIs) along with any injuries sustained by victims are typically recorded in long descriptive narratives.

**Objective:** We aimed to determine if an automated text mining method could identify abuse types and any injuries sustained by domestic violence victims in narratives contained in a large police dataset from the New South Wales Police Force.

**Methods:** We used a training set of 200 recorded domestic violence events to design a knowledge-driven approach based on syntactical patterns in the text and then applied this approach to a large set of police reports.

**Results:** Testing our approach on an evaluation set of 100 domestic violence events provided precision values of 90.2% and 85.0% for abuse type and victim injuries, respectively. In a set of 492,393 domestic violence reports, we found 71.32% (351,178) of events with mentions of the abuse type(s) and more than one-third (177,117 events; 35.97%) contained victim injuries. “Emotional/verbal abuse” (33.46%; 117,488) was the most common abuse type, followed by “punching” (86,322 events; 24.58%) and “property damage” (22.27%; 78,203 events). “Bruising” was the most common form of injury sustained (51,455 events; 29.03%), with “cut/abrasion” (28.93%; 51,284 events) and “red marks/signs” (23.71%; 42,038 events) ranking second and third, respectively.

**Conclusions:** The results suggest that text mining can automatically extract information from police-recorded domestic violence events that can support further public health research into domestic violence, such as examining the relationship of abuse types with victim injuries and of gender and abuse types with risk escalation for victims of domestic violence. Potential also exists for this extracted information to be linked to information on the mental health status.

(*J Med Internet Res* 2019;21(3):e13067) doi: [10.2196/13067](https://doi.org/10.2196/13067)

**KEYWORDS**

domestic violence; injuries; abuse types; text mining; rule-based approach; police narratives

## Introduction

### Background

Domestic violence is a global social and public health phenomenon with important health consequences that affect thousands of lives each year [1-3]. It can be defined as “any incident of threatening behavior, violence (or psychological, physical, sexual, financial, emotional) abuse between adults who are or have been an intimate partner or family member, regardless of gender or sexuality” [4-6]. However, domestic violence can also occur in other relationship structures such as between a caregiver and a dependent person, including a child, or those living together in a household but not in an intimate relationship [4,5]. A multicountry violence study conducted by the World Health Organization estimates a prevalence of 15%-71% in physical and sexual partner violence toward women [1,3]. In Australia, in 2018, one of six women and one of 16 men experienced physical or sexual violence by a current or previous partner [7]. Domestic violence has various forms—from physical to emotional and verbal abuse. The type of abuse received and perpetrated may vary by gender, with each type bearing short- and long-term (physical and mental) health consequences for the victims [8-11]. Domestic violence bears a significant economic cost: Within Australia alone, the cost of violence against women was around Aus \$22.2 billion in 2015-2016 [2,3,12].

The New South Wales Police Force (NSWPF) recorded 123,330 domestic violence-related events in 2017 in WebCOPS (Web Computerised Operational Policing System), a Web-based interface for the COPS, which enables the police to capture and analyze crime information on an organization-wide basis [13]. WebCOPS contains detailed information about domestic violence events as both structured fields (date of birth, Aboriginal status, whether weapons were used, etc) and free unstructured text called “event narratives.” An event can contain more than one text narrative describing, in detail, alleged incident(s) that occurred between the person of interest (POI) and the victim, information regarding the circumstances of the event, and any action(s) taken by the police. Narratives are frequently written without a specific structure, featuring various misspellings, typographical and grammatical errors, and (sometimes informal) acronyms and abbreviations that can have different meanings depending on the context [13].

Domestic violence event narratives contain a wealth of important information regarding injuries and abuse types, which is not found in the medical records unless medical attention is sought, although even attainment of medical attention may not be flagged as related to domestic violence. However, the volume of the recorded data along with the associated long unstructured narratives makes it difficult to identify potentially meaningful information through traditional ethnographic/qualitative research methods involving eyeballing the records. One research paper recently commented that “...there is no systematic way to extract information from these [police] narratives other than by manual review” [14].

### Prior Work

There is a need for methods that can automatically extract information of interest from large volumes of data in a short time. Text mining has been used for more than 30 years to harvest information from unstructured text in many fields, particularly in biomedicine [15-20]. Recent efforts have sought to text mine crime-related information from online media publications [21-23], with limited attempts to process police reports [13,24-28]. Previous work extracted data on the names, narcotic drugs, and weapons with varying degrees of success (F1-score ranging from 46% to 81%) through named entity extraction [24,25] and police report classification of events as domestic violence or nondomestic violence related, using an unsupervised clustering technique that correctly classified 44% of the reports set aside for manual inspection [26]. Other efforts included recognition of crime-related information (such as drugs, weapons, and facial features) from witness narratives through dictionaries and rules, with F1-scores ranging from 82% to 93% [27,28]. Recently, Karystianis et al applied a rule-based approach combined with manually crafted dictionaries to extract mentions of mental illnesses for POIs and victims from police text narratives of recorded domestic violence events with an average F1-score of 84% [13].

### Aim

In this paper, we investigate whether the application of a text mining method can automatically extract abuse types (conducted by POIs) and sustained victim injuries from a large-scale corpus of 492,393 domestic violence events.

## Methods

### Data

We used a corpus of 492,393 domestic violence events provided to the researchers by the NSWPF, occurring from January 2005 to December 2016 [13]. The domestic violence events were flagged in WebCOPS as “domestic violence related,” the description of violence was coded as “domestic,” and the relationship between the victim and the POI included any of the following: “spouse/partner” (including ex-spouse/ex-partner), “boyfriend/girlfriend” (including ex-boyfriend/ex-girlfriend), “parent/guardian” (including step/foster), “child” (including step/foster), “sibling,” “other member of family” (including kin), or “carer.” These events covered the following categories: various types of assaults; breaches of Apprehended Violence Orders; homicides; malicious damage to property; and offense against another person such as intimidation, kidnapping, abduction, and harassment. These data included only events with recorded physical assaults and any cases with stalking, sexual assault, and young POIs were not included.

Permission to access the narratives was granted by the NSWPF following ethics approval from the University of New South Wales Human Research Ethics Committee (Ref: HC16558). Due to the inclusion of sensitive and personal information (eg, name, surname, and address) in the narratives, all processing was undertaken at the NSWPF headquarters. Only de-identified,

extracted outputs were allowed to be taken offsite for further analysis.

We used a total of 300 narratives for training, development (used to enhance the performance of the rules), and evaluation purposes (100 each). These sets are described in more detail in our previous work [13]. A hypothetical de-identified narrative is shown in Figure 1.

### Categorizing Abuse Types

We categorized specific abuse types (ie, details of the abuse behavior) using several sources into nine categories [12,29,30]

with 44 abuse types (Table 1). Although the provided data did not include domestic violence events involving sexual assault and stalking, there were still cases wherein these types of abuse were described in an event. Several nonspecific forms of violence (eg, “bashing,” “smack,” “assaulted,” and “clipping”) were categorized as “assault (unspecified).” A more detailed explanation of the abuse types is provided in Multimedia Appendix 1. A total of 17 common injury types were examined, including scratching, grazing, red mark/sign, tear off (nail), bruising, cut/abrasion, swelling, lump, other, fracture, black eye, broken tooth, burn mark, stab wound, bite mark, soreness, and bleeding.

**Figure 1.** A hypothetical example of a domestic violence event narrative as recorded by the New South Wales Police Force. Blue-highlighted terms indicate the annotated victim injuries, and yellow-highlighted terms indicate the abuse types.

-- Event XXXXXXXXXX : -- Created : XXXXXXXXXXXXXXXXXXXXXXXX -- Narrative 1 of 3 --

The victim named XXX and the defendant, XXX have been living together in XXX, XXX with the victims two children, age XXX and XXX. They have been involved romantically for three years. Due to his frequent drug abuse, the defendant has psychotic episodes in which he physically abuses the victim.

In XXX, around morning, the defendant asked the victim to make breakfast. After the victim went to the kitchen to start preparing breakfast, the defendant snapped and flipped over the dining table screaming “ I told you to make breakfast!”. The victims daughter started crying and the defendant yelled “make that brat quiet, or I will!”. The victim asked the defendant if he has taken his medication to which he replied that he did. Just when the victim was going upstairs to put laundry, the defendant stood in front of her and **slapped** her with his left hand across her face. He then began to **kick** the victim before placing his hands around the victims throat and starting **choking** her.

The victim screamed “ what the hell is wrong with you?”. The defendant then started crying and saying that he was sorry and he did not mean anything. At this point, the victim called her mother and left the house with her two children going immediately at the police station in XXX. The police observed a **large bruise** on the right side of the victims face and **red marks** around her throat.

INJURIES: Victim has a **bruise** on her face and **red marks** in the throat area.

EVIDENCE AND EFFECTS OF ALCOHOL & DRUGS: Nil

MENTAL HEALTH & OTHER ISSUES: The defendant suffers from paranoid schizophrenia and is an alcoholic.

FIREARMS / DANGEROUS WEAPONS: kitchen knife

FEARS HELD BY THE VICTIM: Victim fears that the accused will continue to physically and emotionally abuse her and is afraid that he will hurt her children.

ACTIONS TAKEN BY THE POLICE: Victim statement recorded. The defendant was arrested and charged.

**Table 1.** Categories of abuse along with abuse types.

Abuse category	Abuse type
Physical assault	Assault (unspecified), biting, blocking, choking, ordered dog attack, dragging, elbowing, attempting to set fire to premises, gagging, grabbing, hair pulling, headbutting, head locking, kicking, kneeing, physical restraining, pulling, punching, pushing, scratching, shaking, slapping, spitting, stabbing, victim being thrown around, limb twisting, attempt to harm a victim with an object or weapon, and hitting the victim with an object or weapon
Threat	Intimidation (via body language) or stating explicit threat(s) to physically harm, sexually assault, and self-harm if the victim does not comply
Sexual assault	Sexual assault (eg, rape)
Emotional/verbal abuse	Self-harming when the victim does not comply, yelling profanities, and other emotional/verbal abuse
Stalking	Stalking, harassment, and forced entry
Financial abuse	Financial control (eg, no access to credit card)
Social abuse	Social restriction and prevent/limit child access
Unclassified	Apprehended Domestic Violence Order breach, chasing, lunging, other, and possession of personal effects (eg, phone and car keys)
Property damage	Property damage (ranging from breaking an item to causing damage to a house or vehicle)

## Rule-Based System Development

### Overview

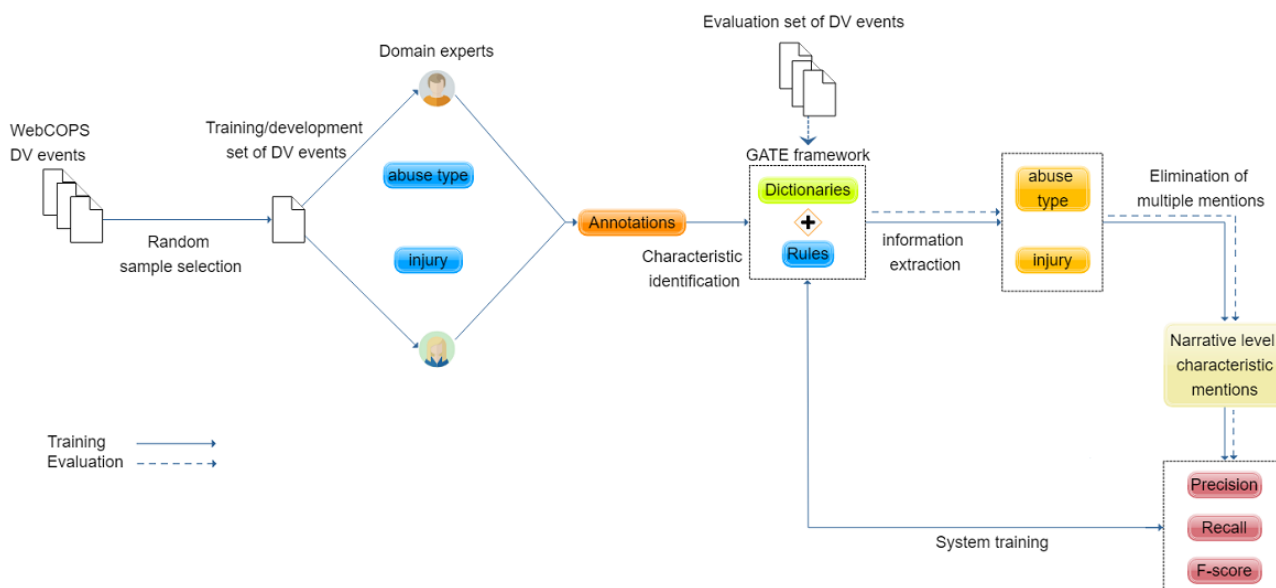
Our method involved the design and implementation of rule-based language expression patterns combined with dictionary terms for the recognition of abuse types and victim injuries at the narrative level. It consisted of the following steps (Figure 2): (1) creation of relevant dictionaries to recognize mentions of abuse types and victim injuries, (2) design and implementation of rules to capture abuse types and victim injuries mentions in context, and (3) aggregation of multiple mentions in each narrative to reach domestic violence event-level annotation.

### Dictionaries

We recognized mentions of task-specific semantic groups through the development of 22 custom-made dictionaries (Table

2). The dictionaries were manually crafted by inspecting the training and the development sets for terms and expressions that describe abuse types (conducted by POIs) and victim injuries, by the first author (GK) and checked by two other authors (AA and PS) to ensure consistency. We used systematic variation (such as plural, past, and present tenses) and also included common misspellings (eg, “stuck” instead of “struck,” “harassment,” and “assalting”) frequently present in the narratives. Although the majority of the terms are noun phrases, for the “threat” dictionary, we included verbal threats made by POIs and manually expanded variations by changing a noun (eg, “your kids are going to have no *father*” to “your kids are going to have no *mother*”) and the surface expressions (“*your dead*” to “*you’re dead*” or “*you are dead*”).

**Figure 2.** An overview of the text-mining methodology used for the identification of abuse types and victim injuries from domestic violence police event narratives. DV: domestic violence; GATE: General Architecture for Engineering; WebCOPS: Web Computerised Operational Policing System.



**Table 2.** The manually crafted dictionaries and their respective size (number of terms included) used to identify abuse types and victim injuries.

Dictionary name	Size	Description	Examples
Anatomy	108	Anatomical parts of the human body in which a victim has been injured by the POI <sup>a</sup>	Chest, leg, head, neck
Assault	18	Verbs that indicate a nonspecific physical attack	Attacked, clipped, smacking, bashing
Attempt	6	Verbs that suggested a physical effort by the POI to harm the victim	Attempted, aimed, trying, tried
Be	4	Conjugations of the verb “be” in the present and past tense	Is, was, were, are
Confiscate	8	Verbs describing a confiscating act by an offender towards a victim	Confiscated, grabbed, snatched, grabbing
Damage	22	Verbs indicating an act of property damage by the POI	Cracked, burned, shattering, ripping
Degree	14	Adjectives describing the victim’s wound	Superficial, extensive, minor, major
Description	59	Terms (mostly adjectives) describing various attributes of an object such as color or type of made material	Yellow, wooden, serving, frying
Family	31	Various nouns indicating the relationship between individuals	Boyfriend, mother, father, cousin
First person threats	123	Threats made by the POI towards a victim	“I will kill you,” “I am going to bury you,” “I will hunt you down and kill you,” “someone is going to kill you”
Force	8	Verbs describing an offender physically restrain a victim	Forcing, pinned, pinning, kept
Location	15	House locations that a DV <sup>b</sup> event occurred at	Toilet, loungeroom, wall, hallway
Number	10	Numbers in words suggesting the number of criminal counts charged at an offender	One, two, four, six
Object	174	Various objects that were broken or used in a DV event	Table leg, cup, rear door, window
POI	18	Terms that describe an offender in a DV event	Defendant, person of interest (offender), offender accused
Premises	6	Terms describing a residence	Unit, terrace, flat, premises
Preposition	44	Various prepositions suggesting the presence of a victim’s injury in an anatomical part	Under left, lower, upper, front
Start	7	Verbs suggesting the initiation or continuation of an action by the offender	Begun, commenced, continuing, started
Trauma	14	Terms indicating a wound caused by a weapon/object used by the offender towards a victim	Wound, cut, trauma, fracture
Victim	19	Terms describing a victim in a DV event	Victim, vic, pinop (short for person in need for protection), pn (short for pinop),
Weapon	155	Objects used to cause harm or threaten to cause harm to a victim by an offender	Army knife, torch, book, shotgun

<sup>a</sup>POI: person of interest.

<sup>b</sup>DV: domestic violence.

## Rules

We based our rules on syntactical patterns identified in the training and development sets, indicating the presence of an abuse type or victim injury. This work follows the same methodology that we previously developed [13]. The syntactical patterns included frozen syntactical expressions as anchors for certain elements built through specific verbs, noun phrases, and prepositions (eg, “commenced to choke”) and semantic

placeholders identifiable through the application of the manually crafted dictionaries (all possible synonyms describing a victim, such as “victim,” “vic,” and “pinop”). We specifically utilized concept enumeration, since it frequently appeared in the training and development sets (eg, “Injuries: *Swollen hand, soreness and scratch under left eye* [mentions of victim’s injuries]”).

General Architecture for Text Engineering (GATE) [31], a text mining framework for annotating and categorizing text, enabling

information recognition, was used to create and apply our rules. The observed syntactical patterns were converted into rules via Java Annotations Pattern Engine, GATE's pattern-matching language. A total of 64 rules were created ([Multimedia Appendix 2](#)).

### Elimination of Multiple Mentions

More than one syntactical pattern may be matched in an event narrative and may refer to one or more mentions of abuse types of victim injuries (that can be duplicates). This led to the extraction of highly variable mentions of abuse types and victim injuries (eg, "punch," "punched," and "punching" are variations of the same abuse type ["punching"]; "bruised," "bruises," and "purple marks" are variations of the same injury ["bruising"]). Each mention is therefore mapped to its "canonical" representative, and only one mention for each abuse type or injury is kept and used to "tag" the domestic violence narrative. For example, if, in a domestic violence event report, we have extracted three mentions of the abuse type "punching" and two mentions of the abuse type "kicking," we only annotate two abuse types—"punching" and "kicking"—at the domestic violence event level.

## Results

### Evaluation

The text mining system was evaluated on a set of 100 previously unseen, randomly chosen domestic violence event reports. The set was manually inspected and annotated by the first and second authors (GK and AA) who identified the type(s) of abuse and victim injuries. The inter-annotator agreement calculated as the absolute agreement rate [32] was 91%, suggesting reliable annotations. Performance of the methodology was evaluated at the narrative level (after eliminating any multiple characteristic mentions). We calculated the precision (the number of true positives against the number of true positives and false positives), recall (the number of true positives against the number of true positives and false negatives), and F1-score (the harmonic mean between precision and recall) at the domestic violence event level using standard definitions [33]. We defined true positive as the detection of a correct mention in an event;

false positive as the extraction of any unrelated mention that has not been annotated manually; false negative as the correct mention that was not detected by our method; and true negative as the case where our method did not identify any mentions when none were annotated.

The results are shown in [Table 3](#). Injuries and abuse types returned F1-scores above 85%, suggesting reliable and consistent results with small but expected drops from the training (5.5% and 9.6%, respectively) and development sets (3.9% and 6.7%, respectively). In particular, the precision was 90.2% for abuse types and 85.0% for the victim injuries, with a small decrease from the development set (2.6% and 5.2%, respectively). In a similar fashion, recall was 89.6% and 86.3% for the abuse types and victim injuries, respectively, with a drop of 5.2% and 8.0%, respectively, when compared to the values of the development set. However, the evaluation set had a significantly smaller number of victim injury mentions (n=66) from the development set (n=88) and the training set (n=83); therefore, its recall value should be considered with caution.

### Large-Scale Corpus Analysis

Given the relatively accurate results of the method in identifying abuse types and victims' injuries, we applied the method to the corpus of 492,393 domestic violence events. Over 71.32% of events (351,178) had an identified abuse type as mentioned in the report, whereas more than one-third of those events (177,607; 36.07%) contained a victim injury ([Table 4](#)).

Of the 44 abuse types, "emotional/verbal abuse" (117,488; 33.46%) was the most common, followed by "punching" (86,322; 24.58%) and "property damage" (78,203; 22.27%). A total of 35.45% (124,498 events) of domestic violence events contained only one identified abuse type, whereas 33.83% (118,819 events) of domestic violence events included three to five different abuse types ([Table 5](#)).

The most frequent injury type was "bruising" (51,455; 29.03%), followed by "cut/abrasion" (51,284; 28.93%) and "red marks/signs" (42,038; 23.71%) ([Table 6](#)). A total of 105,493 domestic violence events (59.56%) had only one form of injury, and 24.48% (43,373) of domestic violence events had two forms of injury ([Table 7](#)).

**Table 3.** Performance of the system on the training, development, and evaluation sets for the identification of abuse types and victim injuries with true positive, false positive, and false negative results.

Set and characteristic	Precision (%)	Recall (%)	F1-score (%)	True positive (%)	False positive (%)	False negative (%)
<b>Evaluation</b>						
Abuse type	90.2	89.6	89.8	259	28	30
Injury	85.0	86.3	85.6	57	10	9
<b>Development</b>						
Abuse type	92.8	94.8	93.7	310	24	17
Injury	90.2	94.3	92.3	83	9	5
<b>Training</b>						
Abuse type	93.9	96.3	95.3	293	19	11
Injury	93.1	97.5	95.2	81	6	2

**Table 4.** Number of domestic violence events containing various abuse types (n=351,178).

Abuse type	Events, n (%)
Assault (unspecified)	171,323 (48.79)
Emotional/verbal abuse	117,488 (33.46)
Punching	86,322 (24.58)
Property damage	78,203 (22.27)
Intimidation	75,662 (21.55)
Grabbing	66,728 (19.00)
Pushing	62,794 (17.88)
Scratching	20,493 (5.84)
Physical restraining	20,014 (5.70)
Kicking	19,435 (5.53)
Slapping	17,474 (4.98)
ADVO <sup>a</sup> breach	16,903 (4.81)
Attempting to hit with an object or weapon	13,592 (3.87)
Hair pulling/dragging by hair	13,048 (3.72)
Choking	11,325 (3.22)
Spitting	9341 (2.66)
Hitting with an object or weapon	8387 (2.39)
Other	7135 (2.03)
Pulling	6373 (1.81)
Victim being thrown around	5255 (1.50)
Lunging	4685 (1.33)
Possession of personal effects	3265 (0.93)
Blocking	3163 (0.90)
Harassment	3100 (0.88)
Stalking	2940 (0.84)
Self-harming	2597 (0.74)
Biting	2285 (0.65)
Dragging	2216 (0.63)
Shaking	2098 (0.60)
Stabbing	1903 (0.54)
Forced entry	1779 (0.51)
Headlocking	1482 (0.42)
Chasing	1324 (0.38)
Kneeing	1321 (0.38)
Gagging	1161 (0.33)
Elbowing	225 (0.06)
Limb twisting	173 (0.05)
Headbutting	148 (0.04)
Sexual assault	125 (0.04)
Prevent child access	91 (0.03)
Social restriction	40 (0.01)
Financial control	29 (0.01)

Abuse type	Events, n (%)
Attempting to set fire to premises	28 (0.01)
Ordered dog attack	1 (0.00)

<sup>a</sup>ADVO: Apprehended Domestic Violence Order.

**Table 5.** Domestic violence events according to the number of abuse types (n=351,178).

Number of abuse type(s)	Events, n (%)
1	124,498 (35.45)
2	89,342 (25.44)
3-5	118,819 (33.83)
6-9	17,951 (5.11)
>10	568 (0.16)
Total	351,178 (100.0)

**Table 6.** Number of events containing various injury types (n=177,607).

Injury type	Events, n (%)
Bruising	51,455 (29.03)
Cut/abrasion	51,284 (28.93)
Red mark(s)	42,038 (23.71)
Swelling	32,581 (18.38)
Soreness	26,729 (15.08)
Other	19,778 (11.16)
Bleeding	19,154 (10.81)
Fracture(s)	17,531 (9.89)
Lump	9482 (5.35)
Grazing	7305 (4.12)
Black eye(s)	2994 (1.69)
Scratching	2399 (1.35)
Bite mark(s)	2350 (1.33)
Stab wound(s)	2346 (1.32)
Burn mark(s)	1382 (0.78)
Broken tooth	620 (0.35)
Tear off nail(s)	7 (0.00)

**Table 7.** Domestic violence events according to the number of victim injury types (n=177,607).

Number of injury types	Events, n (%)
1	105,493 (59.56)
2	43,373 (24.49)
3-4	25,678 (14.49)
5-6	2484 (1.40)
≥7	89 (0.05)
Total	177,117 (100.0)



## Discussion

### Principal Results

To the best of our knowledge, this analysis represents the first attempt to capture domestic violence–related abuse and victim injuries using a large, population-level corpus of domestic violence events recorded by the police. The identification of abuse types conducted by POIs and various injuries sustained by victims in domestic violence disputes are not recorded in the structured information of the WebCOPS database fields. We therefore focused on the narrative part, where the application of our knowledge-driven approach has identified rich information and has the potential to be used for better understanding domestic violence and the development of related prevention interventions, surveillance, and reporting.

Our findings derived from text mining present a more detailed picture of the types of injuries and abuse occurring in domestic violence events. The most common abuse type in our dataset was nonphysical and involved “emotional/verbal abuse,” which is consistent with the recent findings showing that nonphysical abuse types are more prevalent than physical ones [34] and that victims of domestic violence abuse are more likely to sustain certain types of injuries such as cuts and fractures than others [34,35]. Domestic violence can also take myriad physical forms, ranging from victim intimidation to cases where serious and grievous bodily harm is caused by a specific type of abuse (eg, “punching,” “stabbing,” and “choking”), which have both short- and long-term physical and mental health consequences [9-11].

Through the recognition of various abuse types and related victim injuries, potential exists to develop prevention and intervention guidelines by linking this information to diagnostic data held by health services, so that surveillance and monitoring of the victims can be performed. There is also a possibility to track any potential timelines in which the victim was abused. Moreover, the text mining method can be updated on an ongoing basis to monitor trends and inform risk stratification algorithms, which can drive domestic violence–prevention strategies targeting specific groups.

With the inclusion of domestic violence in the WHO’s Sustainable Development Goals, the need for accurate reporting in this area will be necessary [36]. Text mining the police’s domestic violence event narratives is possibly a source of obtaining very nuanced information on this topic including the cause of the event, the potential role of mental illness and substance (ab)use in the event, the types of abuse perpetrated, injuries sustained, weapons used, and information on relationship status. This invaluable information can then be used to target prevention strategies for use by those providing prevention services to particular groups and to identify warning signs for health care providers. A recent report indicated that in Australia, from 2012-2013 to 2013-2014, one woman was killed each week and one man was killed each month as a result of violence from a current or previous partner [7]. Subsequent analyses of this rich information will aim to examine these issues and identify early warning signs of abuse and domestic violence events, which may improve assistance in preventing homicides in domestic violence settings.

### Error Analysis

Although the level of accuracy was acceptable for large-scale analysis to identify trends in domestic violence events, there were still some errors in both abuse types and victim injuries at the level of individual narrative reports. By inspecting the evaluation set, we observed that the system erroneously extracted few instances (five cases) of several POI injuries as victim injuries, since the rules were triggered for the POIs (eg, “*minor grazing to the right shoulder* [false positive for injury] of the POI”). In other instances (4 cases), victim injuries were incorrectly identified when they actually referred to property damage through ambiguous word and syntactical pattern combinations that indicated an injury (eg, “INJURIES/MEDICAL TREATMENT/DAMAGE TO PROPERTY: *Broken table leg* [false positive for victim injury]”). In 12 domestic violence events, when a victim fought back against a POI, any actions by the victim in self-defence were erroneously extracted as an abuse type (eg, “witness stepped in and *grabbed* [false positive for abuse type] the POI and *pinned him to the ground* [false positive for abuse type] until he calmed down” and “...has admitted she physically *pushed him* [false positive for abuse type] back after he *pushed* [true positive for abuse type] into her”). There were few occasions where an abuse type was recognized but had no domestic violence context (eg, “The Accused was closed inside the caged area, where he began *kicking* [false positive for abuse type] at the door and yelling at the police officers...”), while others had not occurred but were likely to happen in the future (eg, “The victim believes if she stayed at the residence she would definitely have been *bashed* [false positive for abuse type] by the accused and possibly *stabbed* [false positive for abuse type]”).

Although we engineered the rules based on generic syntactical patterns that stated victim injuries and abuse types, these rules ignored a limited number of injury mentions, since they were not explicitly stated to have been sustained by the victim (eg, “*redness* [false negative for injury] and *grazes* [false negative for injury] sighted on back, *dried blood* [false negative for injury] on lips”). Some examples (eight cases) were more implicit and required additional inference using some related terms (eg, “the POI *placed his hand in the middle of the victim’s sternum and applied force* [false negative for injury] causing her pain and shortness of breath”). Cases like these were the majority of false negatives for abuse types, suggesting that abuse types such as “grabbing” and “punching” can have quite a few lexical variations in the narratives, which indicate richness of the contexts.

Additionally, injury or abuse type mentions (six cases) that were accompanied by the victim’s surname were excluded from our rule design, since there was no way to determine from the narrative who was the victim or POI without using the structured part of the record (eg, “xxx had a *bleeding nose* [false negative for injury]” and “xxx yelled *verbal abuse* [false negative for abuse type] at her”).

### Limitations

Our text mining system could have missed cases due to more specialized or explicit mentions of abuse types occurring in

domestic violence events, since we based our extraction rules on the information contained in only 200 narratives. Despite incorporation of all types of abuse, there are still likely to be cases in which we probably did not identify explicit types. The relatively smaller number of injury mentions in the evaluation set (when compared to one of the abuse types) could explain the lower performance for the injuries. Nonetheless, we designed our rules based on common syntactical patterns that would attribute abuse types/injury mentions toward POIs and victims, respectively, in order to avoid the generation of false negatives; hence, our recall was higher than the precision in all three datasets. Nevertheless, this approach was able to identify the victim's actions as types of POI's abuse as well as POI's injuries as those of the victim in some instances. This suggests that more specific engineered rules could address this issue. Similarly, although we included the basic and most common forms of injuries, there would be instances containing other causes of injuries or particular abuse types leading to specific injuries that probably have been excluded from our approach. Additionally, the implementation of spell-checking algorithms could assist in the identification of any misspelled abuse types or injuries and potentially elevate performance.

Our analysis of the results from the large corpus of domestic violence events is limited to the abuse types and victim's injuries. We plan to use this information in combination with administrative data collections on mental illness to further examine the nexus between mental illness and domestic violence and explore the relationship of abuse types with gender and

victim injuries. It is pertinent to inquire whether domestic violence victims with mental illness are more vulnerable than those without mental illness in this large-scale dataset spanning 10 years, to identify new intel. Further analysis of the results combined with demographic variables can show interesting aspects of the data in the area of the domestic violence, from prevalence and incidence rates in specific cohorts to risk factors for the occurrence (or recurrence) of domestic violence events. The combination of victim injuries from clinical data resulting from health service contacts could assist in the early identification of victim abuse and the implementation of intervention strategies. Modelling will be used to investigate whether POI characteristics can predict severity of abuse and similarly, whether certain victim phenotypes are prone to particular types of abuse.

## Conclusions

We demonstrated that a knowledge-driven approach can be used for the automated extraction of abuse types and victim injuries involved in domestic violence events. The performance was encouraging, with 90.2% and 85.0% precision for abuse types and injuries, respectively, further implicating that text mining can be used to extract meaningful information from these unstructured data on a large scale. The identified information has enabled us to confirm the magnitude of abuse that victims endure during domestic violence. The results can be used to support further public health research that aims to assess the profiling of POIs involved in domestic violence events and to alter existing intervention policies for victims of abuse.

---

## Acknowledgments

The authors would like to thank the NSWPF for their assistance with this project, particularly Dr Chris Devery, Dr Christie Wallace, John Blanchette, Erin Sharland, and Nicole Grant. This research was supported by an Australian Institute of Criminology/Criminology Research Grant (34/15-16).

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Brief description of the extracted abuse types.

[\[PDF File \(Adobe PDF File\), 215KB-Multimedia Appendix 1\]](#)

---

## Multimedia Appendix 2

Rule examples for recognition of abuse types and victim injuries.

[\[PDF File \(Adobe PDF File\), 192KB-Multimedia Appendix 2\]](#)

---

## References

1. Howard LM, Trevillion K, Khalifeh H, Woodall A, Agnew-Davies R, Feder G. Domestic violence and severe psychiatric disorders: prevalence and interventions. *Psychol Med* 2010 Jun;40(6):881-893. [doi: [10.1017/S0033291709991589](https://doi.org/10.1017/S0033291709991589)] [Medline: [19891808](https://pubmed.ncbi.nlm.nih.gov/19891808/)]
2. Robinson L, Spilsbury K. Systematic review of the perceptions and experiences of accessing health services by adult victims of domestic violence. *Health Soc Care Community* 2008 Jan;16(1):16-30. [doi: [10.1111/j.1365-2524.2007.00721.x](https://doi.org/10.1111/j.1365-2524.2007.00721.x)] [Medline: [18181812](https://pubmed.ncbi.nlm.nih.gov/18181812/)]
3. Trevillion K, Oram S, Feder G, Howard LM. Experiences of domestic violence and mental disorders: a systematic review and meta-analysis. *PLoS One* 2012;7(12):e51740 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0051740](https://doi.org/10.1371/journal.pone.0051740)] [Medline: [23300562](https://pubmed.ncbi.nlm.nih.gov/23300562/)]

4. Home Office Statistical Bulletin. 2008. Crime in England and Wales 2007/08: Findings from the British Crime Survey and police recorded crime URL: [http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/17\\_07\\_08\\_crime\\_statistics\\_200708.pdf](http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/17_07_08_crime_statistics_200708.pdf) [accessed 2019-02-22] [WebCite Cache ID 76MoKgYtI]
5. Briodi A. Sydney City Council and NSWPF. 2010. Domestic Violence is a crime electronic resource URL: <https://catalogue.nla.gov.au/Record/4984779> [accessed 2018-07-11] [WebCite Cache ID 70p4JfaZT]
6. Australian Government - Department of Social Services. The National Plan to Reduce Violence against Women and their Children 2010-2022 URL: <https://www.dss.gov.au/women/programs-services/reducing-violence/the-national-plan-to-reduce-violence-against-women-and-their-children-2010-2022> [accessed 2019-02-28] [WebCite Cache ID 76X39GmnR]
7. Australian Institute of Health and Welfare. 2018. Family, domestic and sexual violence in Australia URL: <https://www.aihw.gov.au/reports/domestic-violence/family-domestic-sexual-violence-in-australia-2018/contents/summary> [accessed 2018-12-07] [WebCite Cache ID 74TboPEi5]
8. Foshee V. Gender differences in adolescent dating abuse prevalence, types and injuries. *Health Educ Res* 1996;11(3):275-286. [doi: [10.1093/her/11.3.275-a](https://doi.org/10.1093/her/11.3.275-a)]
9. Kelly J, Johnson M. Differentiation among types of intimate partner violence: Research update and implications for interventions. *Family Court Review* 2008 Jul;46(3):476-499. [doi: [10.1111/j.1744-1617.2008.00215.x](https://doi.org/10.1111/j.1744-1617.2008.00215.x)]
10. Capaldi D, Shortt J, Kim H, Wilson J, Crosby L, Tucci S. Official incidents of domestic violence: Types, injury, and associations with nonofficial couple aggression. *Violence and Victims* 2009;24(4):502. [Medline: [19694354](https://pubmed.ncbi.nlm.nih.gov/19694354/)]
11. Cleak H, Schofield M, Axelsen L, Bickerdike A. Screening for Partner Violence Among Family Mediation Clients: Differentiating Types of Abuse. *J Interpers Violence* 2018 Apr;33(7):1118-1146. [doi: [10.1177/0886260515614559](https://doi.org/10.1177/0886260515614559)] [Medline: [26681786](https://pubmed.ncbi.nlm.nih.gov/26681786/)]
12. KPMG. 2016. The cost of violence against women and their children in Australia URL: [https://www.dss.gov.au/sites/default/files/documents/08\\_2016/the\\_cost\\_of\\_violence\\_against\\_women\\_and\\_their\\_children\\_in\\_australia\\_-\\_summary\\_report\\_may\\_2016.pdf](https://www.dss.gov.au/sites/default/files/documents/08_2016/the_cost_of_violence_against_women_and_their_children_in_australia_-_summary_report_may_2016.pdf) [accessed 2019-03-01] [WebCite Cache ID 74TbjRYfR]
13. Karystianis G, Adily A, Schofield P, Knight L, Galdon C, Greenberg D, et al. Automatic Extraction of Mental Health Disorders From Domestic Violence Police Narratives: Text Mining Study. *J Med Internet Res* 2018 Sep 13;20(9):e11548. [doi: [10.2196/11548](https://doi.org/10.2196/11548)] [Medline: [30213778](https://pubmed.ncbi.nlm.nih.gov/30213778/)]
14. Macdonald W, Fitzgerald J. NSW Government: Justice – Bureau of Crime Statistics and Research. 2014. Understanding fraud: The nature of fraud offences recorded by NSW Police URL: [https://www.bocsar.nsw.gov.au/Pages/bocsar\\_pages/Fraud.aspx](https://www.bocsar.nsw.gov.au/Pages/bocsar_pages/Fraud.aspx) [accessed 2019-02-22] [WebCite Cache ID 76Moaevm8]
15. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. *Int J Methods Psychiatr Res* 2016 Dec;25(2):86-100. [doi: [10.1002/mpr.1481](https://doi.org/10.1002/mpr.1481)] [Medline: [26184780](https://pubmed.ncbi.nlm.nih.gov/26184780/)]
16. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004;11(5):392-402 [FREE Full text] [doi: [10.1197/jamia.M1552](https://doi.org/10.1197/jamia.M1552)] [Medline: [15187068](https://pubmed.ncbi.nlm.nih.gov/15187068/)]
17. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17(5):507-513 [FREE Full text] [doi: [10.1136/jamia.2009.001560](https://doi.org/10.1136/jamia.2009.001560)] [Medline: [20819853](https://pubmed.ncbi.nlm.nih.gov/20819853/)]
18. Spasić I, Livsey J, Keane JA, Nenadić G. Text mining of cancer-related information: review of current status and future directions. *Int J Med Inform* 2014 Sep;83(9):605-623 [FREE Full text] [doi: [10.1016/j.ijmedinf.2014.06.009](https://doi.org/10.1016/j.ijmedinf.2014.06.009)] [Medline: [25008281](https://pubmed.ncbi.nlm.nih.gov/25008281/)]
19. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. *J Biomed Inform* 2018 Jan;77:34-49 [FREE Full text] [doi: [10.1016/j.jbi.2017.11.011](https://doi.org/10.1016/j.jbi.2017.11.011)] [Medline: [29162496](https://pubmed.ncbi.nlm.nih.gov/29162496/)]
20. Karystianis G, Dehghan A, Kovacevic A, Keane JA, Nenadic G. Using local lexicalized rules to identify heart disease risk factors in clinical notes. *J Biomed Inform* 2015 Dec;58 Suppl:S183-S188 [FREE Full text] [doi: [10.1016/j.jbi.2015.06.013](https://doi.org/10.1016/j.jbi.2015.06.013)] [Medline: [26133479](https://pubmed.ncbi.nlm.nih.gov/26133479/)]
21. Matto G, Mwangoka J. Detecting crime patterns from Swahili newspapers using text mining. *IJKEDM* 2017;4(2):145-156. [doi: [10.1504/IJKEDM.2017.086244](https://doi.org/10.1504/IJKEDM.2017.086244)]
22. Nokhbeh Zaem R, Manoharan M, Yang Y, Barber K. Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security* 2017 Mar;65:50-63. [doi: [10.1016/j.cose.2016.11.002](https://doi.org/10.1016/j.cose.2016.11.002)]
23. Arulanandam R, Savarimuthu B, Purvis M. Extracting crime information from online newspaper articles. 2014 Jan 20 Presented at: Proceedings of the Second Australasian Web Conference-Volume 155; 2014; Auckland, New Zealand.
24. Chau M, Xu J, Chen H. Extracting meaningful entities from police narrative reports. 2002 May 19 Presented at: Proceedings of the annual national conference on Digital government research; 2002; Los Angeles, USA.
25. Ananyan S. AMCIS 2004 Proceedings. 2004. Crime pattern analysis through text mining URL: <https://aisel.aisnet.org/amcis2004/236/> [accessed 2019-03-01] [WebCite Cache ID 76YY4WcQD]
26. Poelmans J, Elzinga P, Viaene S, Dedene G. Formally analysing the concepts of domestic violence. *Expert Systems with Applications* 2011 Apr;38(4):3116-3130. [doi: [10.1016/j.eswa.2010.08.103](https://doi.org/10.1016/j.eswa.2010.08.103)]

27. Ku C, Iriberry A, Leroy G. Crime information extraction from police and witness narrative reports. 2008 Presented at: Technologies for Homeland Security, IEEE Conference on; 2008; Boston, USA.
28. Iriberry A, Leroy G. Natural language processing and e-government: Extracting reusable crime report information. 2007 Presented at: Information Reuse and Integration, IEEE International Conference; 2007; Las Vegas, USA.
29. White Ribbon Australia. 2018. Physical Abuse URL: <https://www.whiteribbon.org.au/understand-domestic-violence/types-of-abuse/physical-abuse/> [WebCite Cache ID 74TbzEdGO]
30. Mouzos J, Makkai T. Australian Institute of Criminology. 2004. Women's experiences of male violence: findings from the Australian component of the International Violence Against Women Survey (IVAWS) URL: <https://aic.gov.au/publications/rpp/rpp56> [accessed 2019-03-01] [WebCite Cache ID 76YYLb70T]
31. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS Comput Biol 2013;9(2):e1002854 [FREE Full text] [doi: [10.1371/journal.pcbi.1002854](https://doi.org/10.1371/journal.pcbi.1002854)] [Medline: [23408875](https://pubmed.ncbi.nlm.nih.gov/23408875/)]
32. Ananiadou S, McNaught J, editors. Text Mining for Biology and Biomedicine. Boston, MA: Artec House Publishers; 2006.
33. Ananiadou S, Kell DB, Tsujii J. Text mining and its potential applications in systems biology. Trends Biotechnol 2006 Dec;24(12):571-579. [doi: [10.1016/j.tibtech.2006.10.002](https://doi.org/10.1016/j.tibtech.2006.10.002)] [Medline: [17045684](https://pubmed.ncbi.nlm.nih.gov/17045684/)]
34. Outlaw M. No One Type of Intimate Partner Abuse: Exploring Physical and Non-Physical Abuse Among Intimate Partners. J Fam Viol 2009 Feb 27;24(4):263-272. [doi: [10.1007/s10896-009-9228-5](https://doi.org/10.1007/s10896-009-9228-5)]
35. Muelleman R, Lenaghan P, Pakieser R. Battered women: injury locations and types. Annals of emergency medicine 1996;28(5):486-492. [Medline: [8909268](https://pubmed.ncbi.nlm.nih.gov/8909268/)]
36. World Health Organization. The Sustainable Development Goals (SDG) and violence prevention: how do they connect? 2018; URL: [https://www.who.int/violence\\_injury\\_prevention/violence/7th\\_milestones\\_meeting/Butchart\\_SDGs\\_and\\_violence\\_prevention.pdf?ua=1](https://www.who.int/violence_injury_prevention/violence/7th_milestones_meeting/Butchart_SDGs_and_violence_prevention.pdf?ua=1) [accessed 2018-12-07] [WebCite Cache ID 74Tc66boY]

## Abbreviations

- ADVO:** Apprehended Domestic Violence Order  
**DV:** domestic violence  
**GATE:** General Architecture for Engineering  
**NSWPF:** New South Wales Police Force  
**POI:** person of interest  
**WebCOPS:** Web Computerised Operational Policing System

*Edited by G Eysenbach; submitted 11.12.18; peer-reviewed by I Spasic, A Davoudi; comments to author 05.01.19; revised version received 31.01.19; accepted 10.02.19; published 12.03.19*

*Please cite as:*

Karystianis G, Adily A, Schofield PW, Greenberg D, Jorm L, Nenadic G, Butler T  
*Automated Analysis of Domestic Violence Police Reports to Explore Abuse Types and Victim Injuries: Text Mining Study*  
*J Med Internet Res* 2019;21(3):e13067  
URL: <http://www.jmir.org/2019/3/e13067/>  
doi: [10.2196/13067](https://doi.org/10.2196/13067)  
PMID: [30860490](https://pubmed.ncbi.nlm.nih.gov/30860490/)

©George Karystianis, Armita Adily, Peter W Schofield, David Greenberg, Louisa Jorm, Goran Nenadic, Tony Butler. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 12.03.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.