Aggression, Escalation, and Other Latent Themes in Legal Intervention Deaths of Non-Hispanic Black and White Men: Results From the 2003–2017 National Violent Death Reporting System

Alina Arseniev-Koehler, MA, Jacob Gates Foster, PhD, Vickie M. Mays, PhD, MSPH, Kai-Wei Chang, PhD, and Susan D. Cochran, PhD, MS

Objectives. To investigate racial/ethnic differences in legal intervention–related deaths using state-of-the-art topic modeling of law enforcement and coroner text summaries drawn from the 2003–2017 US National Violent Death Reporting System (NVDRS).

Methods. Employing advanced topic modeling, we identified 8 topics consistent with dangerousness in death incidents in the NVDRS death narratives written by public health workers (PHWs). Using logistic regression, we then evaluated racial/ethnic differences in PHW-coded variables and narrative topics among 4981 males killed by legal intervention, while adjusting for age, county-level characteristics, and year.

Results. Black, as compared with White, decedents were younger and their deaths were less likely to include PHW-coded mental health or substance use histories, weapon use, or positive toxicology for alcohol or psychoactive drugs, but more likely to include "gangs-as-an-incident-precipitant" coding. Topic modeling revealed less frequent thematic representation of "physical aggression" or "escalation" but more of "gangs or criminal networks" among Black versus White decedents.

Conclusions. While Black males were more likely to be victims of legal intervention deaths, PHW-coded variables in the NVDRS and death narratives suggest lower threat profiles among Black versus similar White decedents. The source of this greater risk remains undetermined. (*Am J Public Health*. 2021;111(S2):S107–S115. https://doi.org/10.2105/AJPH.2021.306312)

In the United States, Black Americans are disproportionately more likely to die in what are labeled legal intervention–related deaths, such as police shootings.[1] Indeed, legal intervention deaths are estimated to be 2.8 times greater among Blacks than among Whites.[2] These deaths occur under a variety of circumstances, including confrontation with police or other legally acting persons during arrests or while in legal confinement. Legal intervention deaths have become increasingly prevalent over the past 2 decades.[3]

Previous research has sought to identify characteristics of legal intervention decedents and the circumstances surrounding their deaths in the hopes of illuminating the basis of the race disparity. For example, it could be hypothesized that victims who pose a clear threat to legal authorities (such as by being armed or acting unpredictably, perhaps because of mental illness or substance use) might be at increased risk for suffering a legal intervention death. If legal intervention incidents

AJPH Supplement 2, 2021, Vol 111, No. S2

involving Black men include more indicators of threat, such as weapon use, then this could explain the greater risk for death. However, previous work finds that Black legal intervention decedents, compared with similar White or Hispanic decedents, are less likely to be armed, exhibit less evidence of mental health difficulties, and are less likely to have positive toxicology for illicit substances or alcohol at time of death.[2–5] Nevertheless, there may be subtle, but unrecognized, clues in descriptions of lethal legal interventions that could clarify unrecognized risk indicators contributing to the greater likelihood of these deaths of Black men.

In the current study, we applied advanced topic modeling methods to illuminate indicators of threat that may be associated with legal intervention deaths among men varying by race/ethnicity. Employing records from the 2003–2017 National Violent Death Reporting System (NVDRS), we used the written summaries of violent deaths in the NVDRS to inductively identify latent descriptors of threat and dangerousness, such as physical aggression. This possibility has not been studied at scale before, to our knowledge. Our goal was to investigate whether unidentified race-related factors play a role in the excess risk for legal intervention deaths seen among Black men.

METHODS

The NVDRS, compiled by the Centers for Disease Control and Prevention (CDC), provides public health surveillance for violent deaths (e.g., suicide, homicide) in the United States.[1] As of 2017, the NVDRS included information on more than 330 000 deaths forwarded by state public health departments from 34 states and the District of Columbia. The death records themselves are created by trained public health workers (PHWs) who use a common, highly articulated CDC coding manual to assign values to multiple variables (e.g., demographics, factors surrounding the death, toxicology results). PHWs also compose 2 brief, narrative summaries describing the circumstances of the death, one drawn from law enforcement reports and the other from medical examiner or coroner reports. We preprocessed these 2 narratives (to standardize abbreviations and correct misspellings) and combined them. We then applied a phrase detector[6] to transform commonly co-occurring words into single tokens (e.g., "African American" became "African_American"). This resulted in 307 249 death narratives in which the decedent was aged 12 years or older at time of death. Of these, 79 062 were homicides, with the great majority being males (n = 63 263).

We then classified male deaths as to legal intervention status. Following Barber et al.,[7] we categorized deaths as legal intervention–related if at least 1 of the following precoded attributes was present in the record:

1. manner of death coded as "legal intervention,"
2. circumstances of the incident involved a "legal intervention" death,
3. victim–suspect relationship coded as "killed by law enforcement,"
4. in the absence of a coded victim–suspect relationship, homicide committed by law enforcement officer in line of duty or a civilian in self-defense,
5. death received legal intervention *International Classification of Diseases, Tenth Revision* (*ICD-10*; Geneva, Switzerland: World Health Organization; 1992) code (Y35.0–Y35.4, Y35.6, Y35.7, and Y89.0, excluding legal executions), or
6. death occurred while victim was "in custody."

This categorization identified 5159 deaths. Finally, we excluded 2 deaths in which the victim was a law enforcement officer and 176 deaths in which the narrative text had fewer than 50 words, as these narratives are unlikely to provide informative contributions to the topic modeling. This exclusion resulted in a final study sample of 4981 deaths. Most (n = 4238; 85%) were by firearms. Further description of the NVDRS data system, data quality, and data validation is available elsewhere.[8]

Study Measures

Public health worker–coded death features. Death records included decedent age at time of death, race/ethnicity, and incident year. We recoded race into 3 categories (non-Hispanic Black, non-Hispanic White, and other). In addition, we used several PHW-coded variables that may reflect dangerousness of encounters: whether the death was precipitated by a serious crime, the decedent used a weapon in the incident, the killing was motivated by or involved gang members, and the decedent had a history of alcohol or substance use problems or a history of mental health difficulties. These 5 variables were coded as yes (1) versus no or unknown (0). Finally, in approximately 70% of cases, the death record included toxicology reports. We recoded these for reported versus not reported: blood alcohol level 0.08 or higher and positive result for any psychoactive drug.

Characteristics of the county where the death occurred. We linked 2 aspects of county-level data to the NVDRS-coded county where the incident occurred: county-level violent crime rate (offenses per 100 000 population) and proportion of non-White county residents.[9]

Narrative-based topics reflecting dangerousness of encounters. To identify latent topics in the narratives, we applied a topic modeling approach. Topic modeling is a machine-learning technique for extracting latent themes from large-scale text data and scoring themes to specific documents.[10–13] A full description of our technique, as implemented in the NVDRS narratives, can be found elsewhere.[14] Briefly, we first used word2vec[15] with Continuous-Bag-of-Words,[6] a word-embedding method, to quantitatively represent the meanings of words in the NVDRS narratives based on their co-occurrences. This method identifies vector representations for words by predicting the presence or absence of words from their various contexts in the corpus. Words sharing more similar contexts—and thus sharing more meaning—tend to have more similar word vectors; similarity between 2 vectors is measured using cosine similarity. We set the dimensionality of word vectors to be 200; we selected this dimensionality by using established quality metrics for word embeddings.[15,16] The resulting embedding matrix had 28 222 rows (1 for each unique word in the narratives) and 200 columns (1 for each component of the 200-dimensional word vectors). This matrix can be thought of as a semantic space, which encodes the meaning of the corpus text.

Next, employing the approach of Arora et al.,[12] we applied a sparse dictionary learning algorithm (K-singular value decomposition, or K-SVD)[17] on this matrix to identify latent topics in the semantic space. We then used several established topic modeling quality metrics to find the best configuration of the model, including coherence of the topics[18] and distinctiveness of topics from one another.[19] These quality measures supported selection of a final model with 225 topics. Adopting methods developed by Arora et al.[10,11] and Arseniev-Koehler et al.,[14] we then assigned each death record a score (0 or 1) for whether the latent topic was present or not in the narrative. This conservative approach assigns a score of 1 to narratives with any amount of a topic present. Finally, we confirmed the robustness and validity of the topic model by employing narratives' topic assignments to predict, using a logistic regression classifier, 25 PHW-coded binary death characteristics with 10-fold cross validation (e.g., whether the victim died in an emergency department, a firearm was the primary weapon, and the victim was recently released from any kind of institution). Across the 25 outcomes evaluated, mean classification accuracy was 82.3% (SD = 7.4%) on held-out validation folds.

The face validity of these latent topics can be discerned by examining words loading highest on each topic, as indexed by cosine similarity scores (these are the most representative words for the topic). As might be expected, the 225 topics ranged across the descriptive landscape of violent death, including administrative words, weapons and ammunition, sounds and smells, emotions, syntactic jargon, body parts, drugs, and locations. From these, 2 of the authors (S. D. C. and A. A.-K.) independently identified 8 topics with 100% agreement that, by face validity, address aspects of dangerous encounters (Table 1).

Data Analysis

We analyzed data by using multiple imputation procedures as implemented in R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). Missing data were rare (< 2.0% for any variable) except for toxicology reports, which were analyzed as subsamples. As a first step, we used linear regression to investigate racial/ethnic differences in study covariates: age, incident year, and county-level characteristics (crime rate and proportion of non-White residents). Previous work suggests that younger individuals are at heightened risk of death by police force[20] and, given the changing state composition of the NVDRS, we adjusted for effects attributable to this factor. County-level violent crime rates and proportion of non-White residents were also treated as possible confounders given previous findings[21] that legal intervention deaths and racial disparities in policing are linked to the place where the incident happened, not merely characteristics of the victim or encounter.

Next, we examined potential racial/ethnic variation in the word count of summaries by using a negative binomial regression model, adjusting for incident year to account for any potential changes in NVDRS reporting. Last, we used logistic regression to investigate race/ethnicity differences in selected PHW-coded variables and the 8 narrative topics potentially indexing dangerousness of the lethal encounter while adjusting for age, incident year, and county-level factors. In the analyses of narrative topics, we additionally adjusted for the number of words in each narrative; our descriptive analyses found Black decedents' narratives were often shorter in length, and shorter narratives are less likely to include any particular topic. From these analyses, we report adjusted odds ratios and their 95% confidence intervals. We also report results of regression modeling and the Wald F test. All significance tests were based on the criterion of a *P* value of less than .05.

RESULTS

As shown in Table 2, compared with White decedents, Black decedents were

TABLE 1— Latent Topics Within the 2003–2017 National Violent Death Reporting System Legal Intervention Death Narratives Indicative of Dangerous Encounters: United States

Narrative Topic Label	% (SE) of Narratives With This Topic	Ten Most Representative Terms
Physical aggression	84.5 (0.5)	tackled, lunged_toward, began_attacking, advanced_toward, attacked, slapped, intervened, shoved, lunged, pepper_sprayed
Fight beginnings	87.8 (0.5)	fight_ensued, gunfire_erupted, physical_altercation_ensued, another_individual, pistol_whipped, gunman, struggle_ensued, scuffle_ensued, suspect, intruders
Justification	24.4 (0.6)	ruled_justifiable, remains_unsolved, 558, gang_motivated, pedestrian_vs_train, road_rage, random_violence, justifiable_self_defense, 3289, considered_justifiable
Escalation	8.3 (0.4)	becoming_increasingly, becoming_more, become_more, increasingly, noticeably, notably, become_increasingly, profoundly, grown_increasingly, generally
Physical posture	41.5 (0.7)	crouching, silhouette, kneeling, northeast, crouched, walkway, platform, stagger, laying, leaning
Causal language	12.3 (0.5)	sparked, preceded, triggered, precipitated, led, prompted, culminated, may_have_contributed, occurred, completely_unexpected
Gangs or criminal networks	45.8 (0.7)	gang, rival_gang, bloods, crips, gang_activity, drug_trade, crips_gang, rival, rival_gang_members, revenge
Hostile confrontation	39.2 (0.7)	home_invasion_robbery, card_game, drug_transaction, gunfight, shootout, scuffle, hostage_situation, verbal_exchange, confrontation, brawl

Note. The sample size was n = 4981. "Most representative terms" are the 10 terms with highest cosine similarity to the relevant topic.

AJPH Supplement 2, 2021, Vol 111, No. S2

far younger (b = −9.30; $P < .001$) as were decedents of other races/ethnicities (b = −6.89; $P < .001$). More than 69% of all Black men dying by legal intervention were younger than 35 years, and 25% were younger than 22 years. By contrast, just 38% of White men were younger than 35 years with 6% younger than 22 years. Timing of the deaths varied in the NVDRS; Black decedents tended to have an earlier incident year in the database (b = −0.27; $P < .05$) as compared with Whites, while decedents of other races/ethnicities tended to have a more recent incident year (b = 0.51; $P < .01$). We also observed racial/ethnic differences in the county characteristics where the lethal event occurred: compared with Whites, Black decedents tended to die in counties with higher rates of violent crime (b = 249.02; $P < .001$) and higher percentages of non-White residents (b = 19.07; $P < .001$), as did decedents of other races/ethnicities (b = 115.78; $P < .001$ and b = 7.26; $P < .001$, respectively). Narratives describing events for Black decedents included fewer words (adjusted b = −0.28; $P < .001$) than those for White decedents, while narratives for decedents of other races/ethnicities included more words (adjusted b = 0.10; $P < .001$), after we controlled for year when the death occurred.

Legal intervention deaths often occurred in the context of the commission of a serious crime where weapons were used during the encounter (Table 2). Many decedents were also potentially impaired by consumption of alcohol or other psychoactive substances or had histories of mental health or substance use problems. However, compared with death records of White decedents, Black decedents' records less frequently included positive histories of mental health or substance use problems or use of a weapon by the decedent during the incident (Table 3), after we adjusted for covariates. Furthermore, in the subsamples of deaths in which toxicology results were available, Black decedents were less likely than Whites to be intoxicated or to screen positive for psychoactive drugs at the time of death, after we adjusted for covariates.

Narrative summaries of these deaths revealed subtle but critical Black–White differences in how these legal intervention deaths were described by the PHW. Compared with incidents involving White decedents, Black decedents' incidents were less likely to include topics describing physically aggressive actions or indicators of escalation, but more likely to include language related to characterizations of gangs or criminal networks. Finally, compared with White decedents, narratives for decedents of

TABLE 2— Characteristics of Legal Intervention–Related Deaths Among Males, Aged 12 Years and Older, by Racial/Ethnic Background: United States, National Violent Death Reporting System, 2003–2017

Characteristics	Non-Hispanic Black, No. or % (SE)	Non-Hispanic White, No. or % (SE)	Other, No. or % (SE)
No. of deaths	1752	2317	912
County of incident			
County violent crime offenses, no. per 100 000 population***	589.4 (7.9)	340.4 (4.7)	456.2 (7.4)
County non-White population, %**	40.3 (0.4)	21.2 (0.3)	28.5 (0.5)
Victim age in years, mean***	31.3 (0.3)	40.6 (0.3)	33.7 (0.4)
Proximal factors, %			
Positive history alcohol or substance use problems***	7.1 (0.6)	23.9 (0.9)	25.5 (1.4)
Positive history of mental health problems***	6.2 (0.6)	17.8 (0.8)	11.1 (1.0)
Death precipitated by a serious crime*	73.7 (1.1)	74.6 (0.9)	78.2 (1.4)
Victim used a weapon***	52.7 (1.2)	62.4 (1.0)	61.4 (1.6)
Death was motivated by or involved gangs***	3.5 (0.4)	0.9 (0.2)	4.1 (0.7)
Narrative topics for dangerous encounters present, %			
Physical aggression***	78.5 (1.0)	88.0 (0.7)	87.1 (1.1)
Fight beginnings*	87.4 (0.8)	87.0 (0.7)	90.5 (1.0)
Justification	25.4 (1.0)	24.4 (0.9)	22.7 (1.4)
Escalation***	5.0 (0.5)	10.5 (0.6)	9.0 (0.9)
Physical posture, %***	38.0 (1.2)	43.7 (1.0)	42.6 (1.6)
Causal language**	11.7 (0.8)	13.7 (0.7)	10.0 (1.0)
Gangs or criminal networks***	46.8 (1.2)	42.5 (1.0)	52.2 (1.7)
Hostile confrontation***	32.8 (1.1)	41.5 (1.0)	45.3 (1.6)
Characteristic of narrative: narrative word count, mean***	203.9 (3.7)	267.0 (4.5)	310.5 (9.8)
Subsamples of deaths that included toxicology reports			
Blood alcohol level			
No. of deaths tested for alcohol levels	1117	1656	724
Blood alcohol level > 0.8, %***	21.5 (1.0)	35.5 (1.0)	28.5 (1.5)
Presence of psychoactive drugs			
No. of deaths screened for psychoactive drug toxicology	1225	1716	758
Positive for psychoactive drugs, %***	48.7 (1.2)	58.6 (1.0)	61.8 (1.6)

Note. The sample size was 4981. For death records that included toxicology reports, n = 3497 for blood alcohol level and n = 3699 for toxicology screen for psychoactive drugs.

*$P < .05$; **$P < .01$; ***$P < .001$.

other races/ethnicities were also more likely to include the gangs or criminal networks topic.

DISCUSSION

Legal intervention deaths are a relatively rare occurrence among the nation's total homicides.[2,3,7] They are, however, among some of the most highly contested homicides. Investigations of these deaths have revealed that approximately a fifth likely involve mental health crises or substance use problems,[22] which may lead to confrontational behavior with law enforcement (so-called "suicide by cop"). Many others occur in the context of serious, felonious crime. Legal intervention deaths are also marked by important risk patterns wherein Black men, and young Black men especially, are at much greater risk in their interactions with law enforcement than are White

TABLE 3— Partial Results of Logistic Regression Models Estimating Proximal Factors and Narrative Topics Associated With Legal Intervention Deaths Among Males, Age 12 Years and Older, by Race/Ethnicity: United States, National Violent Death Reporting System, 2003–2017

	Race/Ethnicity[a]	
Characteristics	**Non-Hispanic Black, No. or AOR (95% CI)**	**Other, No. or AOR (95% CI)**
Total sample	1752	912
Proximal factors		
Positive history alcohol or substance use problems	0.29 (0.28, 0.37)	1.09 (0.90, 1.33)
Positive history of mental health problems	0.38 (0.30, 0.50)	0.60 (0.47, 0.77)
Death precipitated by a serious crime	0.86 (0.72, 1.01)	1.10 (0.91, 1.33)
Victim used a weapon	0.64 (0.55, 0.74)	0.93 (0.80, 1.11)
Death was motivated by or involved gangs	4.07 (2.32, 7.15)	4.23 (2.40, 7.45)
Narrative topics for dangerous encounters present		
Physical aggression	0.65 (0.52, 0.81)	0.87 (0.67, 1.13)
Fight beginnings	1.11 (0.89, 1.40)	1.20 (0.91, 1.57)
Justification	1.07 (0.90, 1.28)	0.76 (0.63, 0.93)
Escalation	0.73 (0.54, 0.99)	0.78 (0.58, 1.05)
Physical posture	1.01 (0.86, 1.19)	0.79 (0.66, 0.94)
Causal language	1.16 (0.92, 1.45)	0.63 (0.48, 0.82)
Gangs or criminal networks	1.34 (1.14, 1.56)	1.26 (1.07, 1.50)
Hostile confrontation	0.91 (0.78, 1.07)	1.14 (0.97, 1.35)
Subsamples of deaths that included toxicology reports		
Blood alcohol level		
No. of deaths tested for blood alcohol levels	1117	724
Blood alcohol level > 0.8	0.55 (0.45, 0.67)	0.77 (0.63, 0.94)
Presence of psychoactive drugs		
No. of deaths screened for psychoactive drug toxicology	1225	758
Positive for psychoactive drugs	0.60 (0.50, 0.81)	0.93 (0.77, 1.12)

Note. AOR = adjusted odds ratio; CI = confidence interval. The sample size was n = 4981, including 2317 non-Hispanic Whites. For toxicology screening variables, the sample size was n = 3497 (including 1656 Whites) for blood alcohol levels and 3699 (including 1716 Whites) for psychoactive drug toxicology reports. Effects tested by adjusting for decedent's age, characteristics of the county where the incident occurred, year of the incident, and, for narrative topics only, narrative word count.

[a]Referent = non-Hispanic White.

men.[2] Importantly, some of these risk patterns may represent public health opportunities to reduce mortality, particularly among young Black men.

The NVDRS represents an important official source for much of what is known about violent death in the United States, including legal intervention deaths.[1,23] Using variables coded into the NVDRS (e.g., sex, *ICD* code, type of weapon, toxicology report), previous studies have shown that Black men who die by legal intervention appear to pose lower immediate threat to law enforcement as compared with White men who also die by legal intervention.[24] In addition, their deaths are less consistent with possible "suicide by cop" motivations,[2,24] as Black legal intervention decedents are less likely than similar Whites to have histories of mental health problems or substance abuse, or evidence of current drug or alcohol consumption at time of death.[24]

Our study extends these findings by making use of a large, mostly untapped source of information on the legal intervention deaths within the NVDRS death narratives. Traditional qualitative methods are impractical for analyzing the hundreds of thousands of NVDRS narratives, though previously some have mined small portions of them. For example, researchers have used simple text searches[25,26] and case sampling employing combinations of precoded variables(s)[22,24,27,28] to isolate samples of death narratives that are then subjected to traditional qualitative methods. This

work has successfully isolated the importance of bullying in sexual minority youth suicides,[25] characteristics of gang-related killings,[28] and the variables that we used to identify legal intervention deaths in the NVDRS.[7] But much of the potential value of the narratives has remained out of reach.

To that end, we applied a novel machine-learning approach to unlock the thematic information contained within the narratives. Focusing on themes relevant to aggression, threat, and danger posed by the incident, we found that nuanced themes of danger are described in these narratives, such as specific physically aggressive actions (e.g., "lunged toward"). We also observed a pattern of racial/ethnic differences in the frequency of these themes among legal intervention deaths. For example, incidents involving Black male decedents, as compared with similar Whites, were less likely to involve descriptions of physically aggressive actions or escalation (e.g., "becoming increasingly") but more likely to include descriptions of gangs and criminal networks. By using topic modeling, we were able to both confirm and extend characterization of race differences associated with descriptions of legal intervention deaths.

At a more general level, our topic modeling methodology also offers 2 potential benefits when applied to administrative data such as the NVDRS. First, it may prove a useful approach to inductively identify characteristics of death incidents mentioned in the NVDRS narratives, but not as yet included in its precoded variables. Second, it offers a way to validate PHW coding of variables. For example, although gang-motivated or -involved factors, as indexed by the PHW-coded variable, were rarely present in legal intervention deaths, topic modeling revealed that this characterization was quite common in the death narratives. Elsewhere,[28] NVDRS underreporting of gang involvement, especially among deaths involving non-Hispanic Whites and American Indian/Alaska Natives, has already received note. Other topics picked up by our model but not examined in this study (e.g., topics about mental health) might also be useful to extend or triangulate other PHW-coded variables.

Limitations

Our study has several limitations to be considered. First, while we investigated characteristics of the incident, victim, and county where the incident occurred, we did not directly account for the local racial climate, something that has been suggested as playing a role in the racial disparities observed in legal intervention deaths.[29] In some neighborhoods, Black men may be more likely to be stopped by police in the first place, and Blacks are more likely than Whites to be subject to use of force by law enforcement.[2] Second, the NVDRS is a relatively new administrative database that has been shown to include a fairly comprehensive count of fatal police shootings[30] and legal intervention deaths,[7] but only within the limited set of states that it covers. The recent expansion of the NVDRS to include all 50 states will enable more detailed future work on geographic variation.

Third, the NVDRS is a set of abstracted records compiled by multiple state-based PHWs. Although coders undergo extensive training and use a detailed, standardized codebook, and the abstracted information undergoes numerous, ongoing data validation checks, PHWs may still vary somewhat in their reporting. Also, the database summarizes existing death records but does not investigate original circumstances. Hence, data quality may vary in indeterminate ways at multiple levels (the death is underinvestigated, the report is inaccurately or underreported, or the PHW errs in completing the NVDRS record). All of these factors may introduce biases. Finally, our topic modeling approach, though ideal for processing large databases and avoiding limitations of the need for a priori text search terms, uses alternate validity checks rather than traditional text search strategies[25,30] where all identified cases are closely read for accuracy by human analysts. Metrics reported here support the validity of our approach, but some records remain inaccurately classified. The likely direction of bias is toward the null.

Despite these limitations, results presented here provide further evidence for the role of racially infused perceptions of threat in incidents of legal intervention and police shooting deaths.[2] In both the PHW-coded variables and the topics identified via a novel topic modeling approach, our results suggest that circumstances involving Black decedents from legal interventions include fewer indicators of threat when compared with similar White male decedents. These findings hint at differential, race-related threat thresholds for lethal legal intervention. Elsewhere,[31] research indicates that there is a greater propensity to associate perceptions of threat and criminality with Black men where, for example, young Black men's size and muscularity tend to be overestimated. Furthermore, the overestimation occurs more often with Black men than with White men. This propensity appears to lead to overestimation of danger in situations in which law enforcement is required to interpret ambiguous behavior and circumstances.[32] As a consequence, evidence reported elsewhere suggests that Black

individuals are more likely than Whites to be erroneously perceived as holding weapons, when they are in fact holding benign objects.[33] Simulation studies also find that the decision to shoot occurs faster when the target is Black versus White.[29]

Finally, while our core interest lay in characterizing narrative descriptions of the danger or threat posed by the incident, we unexpectedly observed that legal intervention narratives tended to be systematically shorter for Black versus White men. Whether this reflects the amount of information gathered in original law enforcement, coroner, and medical examiner reports or the level of detail retained by PHWs is indeterminable. Regardless of the origin for difference in length, the NVDRS is an official source of information for ascertaining circumstances of death, and shorter narratives provide less information for investigations.

Public Health Implications

Substantively, this work builds on a growing body of scholarship seeking to understand racial disparities in legal intervention deaths, especially police shootings. Using written summaries from public health workers in the 2003–2017 NVDRS, the current study offers new evidence that incidents involving Black decedents who die by legal intervention are described differently than those involving White men. As predicted, the narratives suggest that Black legal intervention deaths, as compared with Whites, are associated with characteristics that pose lower objective threat profiles for law enforcement. At the same time, there is also evidence that these deaths are less fully described for reasons that are unknowable at this point. By applying a state-of-the-art topic modeling technique to a public health administrative database, we were also able to demonstrate the utility and efficiency of this approach, offering a pathway for greater exploration of racial inequities using this important federal data system. Future public health efforts to reduce racial differences in perceived threat or harm in police interactions with young Black males are clearly warranted. AJPH

ABOUT THE AUTHORS

Alina Arseniev-Koehler and Jacob Gates Foster are with the Department of Sociology, University of California Los Angeles (UCLA). Vickie M. Mays is with the Department of Psychology, UCLA, and the Department of Health Policy and Management, UCLA Fielding School of Public Health, and the BRITE Center, UCLA. Kai-Wei Chang is with the Department of Computer Science, UCLA Samueli School of Engineering. Susan D. Cochran is with the Department of Epidemiology, UCLA Fielding School of Public Health, and Department of Statistics, UCLA. Vickie M. Mays and Susan D. Cochran are also Guest Editors for this supplement issue.

CORRESPONDENCE

Correspondence should be sent to Susan D. Cochran, PhD, Department of Epidemiology, UCLA Fielding School of Public Health, 650 Charles E Young South, Los Angeles, CA 90095-1772 (e-mail: cochran@ucla.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

PUBLICATION INFORMATION

Full Citation: Arseniev-Koehler A, Foster JG, Mays VM, Chang K-W, Cochran SD. Aggression, escalation, and other latent themes in legal intervention deaths of non-Hispanic Black and White men: results from the 2003–2017 National Violent Death Reporting System. *Am J Public Health*. 2021;111(S2):S83–S91.

Acceptance Date: March 27, 2021.

DOI: https://doi.org/10.2105/AJPH.2021.306312

CONTRIBUTORS

All authors have contributed substantially to this article. A. Arseniev-Koehler and S. D. Cochran conceptualized and designed the study. All authors contributed to the conceptualization of the topic modeling approach. The analytic plan was developed by A. Arseniev-Koehler, J. G. Foster, K. W. Chang, and S. D. Cochran. A. Arseniev-Koehler implemented analyses along with J. G. Foster and K. W. Chang. The initial draft of the article was written by A. Arseniev-Koehler and S. D. Cochran, and all authors edited subsequent versions of the article.

ACKNOWLEDGMENTS

The authors acknowledge funding from the National Institute of Minority Health and Health Disparities (MD 006923), the National Institute of Mental Health (MH 115344), the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1650604, and the UCLA Bedari Kindness Institute. J. G. Foster is supported by an Infosys Membership in the School of Social Science at the Institute for Advanced Study.

CONFLICTS OF INTEREST

The authors have no conflicts to disclose.

HUMAN PARTICIPANT PROTECTION

This study was ruled exempt from human participant review by the UCLA institutional review board because it used secondary, de-identified data.

REFERENCES

1. Ertl A, Sheats K, Petrovsky E, Betz C, Yuan K, Fowler K. Surveillance for violent deaths—National Violent Death Reporting System, 32 states, 2016. *MMWR Surveill Summ*. 2019;68(9):1–36. https://doi.org/10.15585/mmwr.ss.6809a1
2. DeGue S, Fowler S, Calkins C. Deaths due to use of lethal force by law enforcement: findings from the National Violent Death Reporting System, 17 US states, 2009–2012. *Am J Prev Med*. 2016;51(5):S173–S187. https://doi.org/10.1016/j.amepre.2016.08.027
3. Drowos J, Hennekens C, Levine R. Variations in mortality from legal intervention in the United States—1999 to 2013. *Prev Med*. 2015;81:290–293. https://doi.org/10.1016/j.ypmed.2015.09.012
4. Buehler JW. Racial/ethnic disparities in the use of lethal force by US police. *Am J Public Health*. 2017;107(2):295–297. https://doi.org/10.2105/AJPH.2016.303575
5. Nix J, Campbell B, Byers E, Alpert G. A bird's eye view of civilians killed by police in 2015. *Criminol Public Policy*. 2017;16(1):309–340. https://doi.org/10.1111/1745-9133.12269
6. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. In: *Proceedings of the Language Resources and Evaluation Conference 2010 Workshop on New Challenges for Natural Language Processing Frameworks*. Paris, France: European Language Resources Association; 2010:45–50.
7. Barber C, Azrael D, Cohen D. Homicides by police: comparing counts from the National Violent Death Reporting System, Vital Statistics, and Supplementary Homicide Reports. *Am J Public Health*. 2016;106(5):922–927. https://doi.org/10.2105/AJPH.2016.303074
8. Paulozzi LJ, Mercy J, Frazier L, Annest JL. National Violent Death Reporting System: background and methodology. *Inj Prev*. 2004;10(1):47–52. https://doi.org/10.1136/ip.2003.003434
9. Agency for Healthcare Research and Quality. Social Determinants of Health Database. 2020. Available at: https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html. Accessed March 1, 2021.
10. Arora S, Liang Y, Ma T. A simple but tough-to-beat baseline for sentence embeddings. In: *Proceedings of the 5th International Conference on Learning Representations*. La Jolla, CA: International Conference on Learning Representations; 2017:1–11.

11. Arora S, Li Y, Yingyu L, Ma T, Risteski A. A latent variable model approach to pmi-based word embeddings. *Trans Assoc Comput Linguist*. 2016;4:385–399. https://doi.org/10.1162/tacl_a_00106
12. Arora S, Li Y, Liang Y, Tengyu M, Risteski A. Linear algebraic structure of word senses, with applications to polysemy. *Trans Assoc Comput Linguist*. 2018;6:483–495. https://doi.org/10.1162/tacl_a_00034
13. Ethayarajh K. Unsupervised random walk sentence embeddings: a strong but simple baseline. In: *Proceedings of The Third Workshop on Representation Learning for Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics; 2018:91–100. https://doi.org/10.18653/v1/W18-3012
14. Arseniev-Koehler A, Cochran S, Mays V, Chang K-W, Foster JG. Discourses of death: extracting the semantic structure of lethal violence with machine learning. Preprint. Posted online August 3, 2020:1–32. *SocArXiv*. https://doi.org/10.31235/osf.io/nkyaq
15. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: *Proceedings of the 26th International Conference on Advances in Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc; 2013:3111–3119.
16. Finkelstein L, Gabrilovich E, Matias Y, et al. Placing search in context: the concept revisited. *ACM Trans Inf Syst*. 2002;20(1):116–131. https://doi.org/10.1145/503104.503110
17. Aharon M, Elad M, Bruckstein AK-SVD. An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process*. 2006;54(11):4311–4322. https://doi.org/10.1109/TSP.2006.881199
18. Aletras N, Stevenson M. Evaluating topic coherence using distributional semantics. In: *Proceedings of the 10th International Conference on Computational Semantics*. Potsdam, Germany: Association for Computational Linguistics; 2013:13–22.
19. Dieng A, Ruiz F, Blei D. Topic modeling in embedding spaces. *Trans Assoc Comput Linguist*. 2020;8:439–453. https://doi.org/10.1162/tacl_a_00325
20. Edwards F, Lee H, Esposito M. Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. *Proc Natl Acad Sci U S A*. 2019;116(34):16793–16798. https://doi.org/10.1073/pnas.1821204116
21. Siegel M, Poulson M, Sangar R, Jonathan J. The interaction of race and place: predictors of fatal police shootings of Black victims at the incident, census tract, city, and state levels, 2013–2018. *Race Soc Probl*. 2021; epub ahead of print January 13, 2021. https://doi.org/10.1007/s12552-020-09307-y
22. Lord V. Police responses in officer-involved violent deaths: comparison of suicide by cop and non-suicide by cop incidents. *Police Q*. 2014;17(1):79–100. https://doi.org/10.1177/1098611114522040
23. Nazarov O, Guan J, Chihuri S, Li G. Research utility of the National Violent Death Reporting System: a scoping review. *Inj Epidemiol*. 2019;6(1):18. https://doi.org/10.1186/s40621-019-0196-9
24. Wertz J, Azrael D, Berrigan J, et al. A typology of civilians shot and killed by US police: a latent class analysis of firearm legal intervention homicide in the 2014–2015 National Violent Death Reporting System. *J Urban Health*. 2020;97(3):317–328. https://doi.org/10.1007/s11524-020-00430-0
25. Clark KA, Cochran S, Maiolatesi A, Pachankis J. Prevalence of bullying among youth classified as LGBTQ who died by suicide as reported in the National Violent Death Reporting System, 2003–2017. *JAMA Pediatr*. 2020;174(12):1211–1213. https://doi.org/10.1001/jamapediatrics.2020.0940.
26. Lyons BH, Walters M, Jack S, Petrovsky E, Blair J, Ivey-Stephenson A. Suicides among lesbian and gay male individuals: findings from the National Violent Death Reporting System. *Am J Prev Med*. 2019;56(4):512–521. https://doi.org/10.1016/j.amepre.2018.11.012
27. McNally MR, Patton C, Fremouw W. Mining for murder–suicide: an approach to identifying cases of murder–suicide in the National Violent Death Reporting System restricted access database. *J Forensic Sci*. 2016;61(1):245–248. https://doi.org/10.1111/1556-4029.12887
28. Frazier L Jr, Ortega L, Patel N, Barnes J, Crosby A, Hempstead K. Methods and findings from the National Violent Death Reporting System for identifying gang-like homicides, 2005–2008. *J Natl Med Assoc*. 2017;109(4):272–278. https://doi.org/10.1016/j.jnma.2017.03.001
29. Mesic A, Franklin L, Cansever A, et al. The relationship between structural racism and Black–White disparities in fatal police shootings at the state level. *J Natl Med Assoc*. 2018;110(2):106–116. https://doi.org/10.1016/j.jnma.2017.12.002
30. Conner A, Azrael D, Lyons V, Barber C, Miller M. Validating the National Violent Death Reporting System as a source of data on fatal shootings of civilians by law enforcement officers. *Am J Public Health*. 2019;109(4):578–584. https://doi.org/10.2105/AJPH.2018.304904
31. Wilson JP, Hugenberg K, Rule N. Racial bias in judgments of physical size and formidability: from size to threat. *J Pers Soc Psychol*. 2017;113(1):59–80. https://doi.org/10.1037/pspi0000092
32. Correll J, Park B, Judd C, Wittenbrink B. The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. *J Pers Soc Psychol*. 2002;83(6):1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314
33. Payne BK. Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *J Pers Soc Psychol*. 2001;81(2):181–192. https://doi.org/10.1037/0022-3514.81.2.181