

Utility of open information sources to develop estimates of selected crime and justice indicators:

Measuring arrest-related deaths

Duren Banks, Jason Nance, Michael Planty, Erin Kennedy & Peter Baumgartner

RTI International

Anthony Whyde, Kevin Scott & Connor Brooks

Bureau of Justice Statistics



Federal Committee on Statistical Methodology
2018 Research and Policy Conference

Disclaimer

The views, opinions, and content expressed in this presentation do not necessarily reflect the views, opinions, or policies of the Bureau of Justice Statistics (BJS), the Office of Justice Programs (OJP), or the U.S. Department of Justice (DOJ).

Presentation Overview

- Coverage issues related to traditional methods that rely on agency self-reports
- Arrest-Related Deaths (ARD) pilot study
- Use of open information sources
 - Coding and classification pipeline methodology
 - Efficiency metrics
- Agency survey methodology
- Discussion

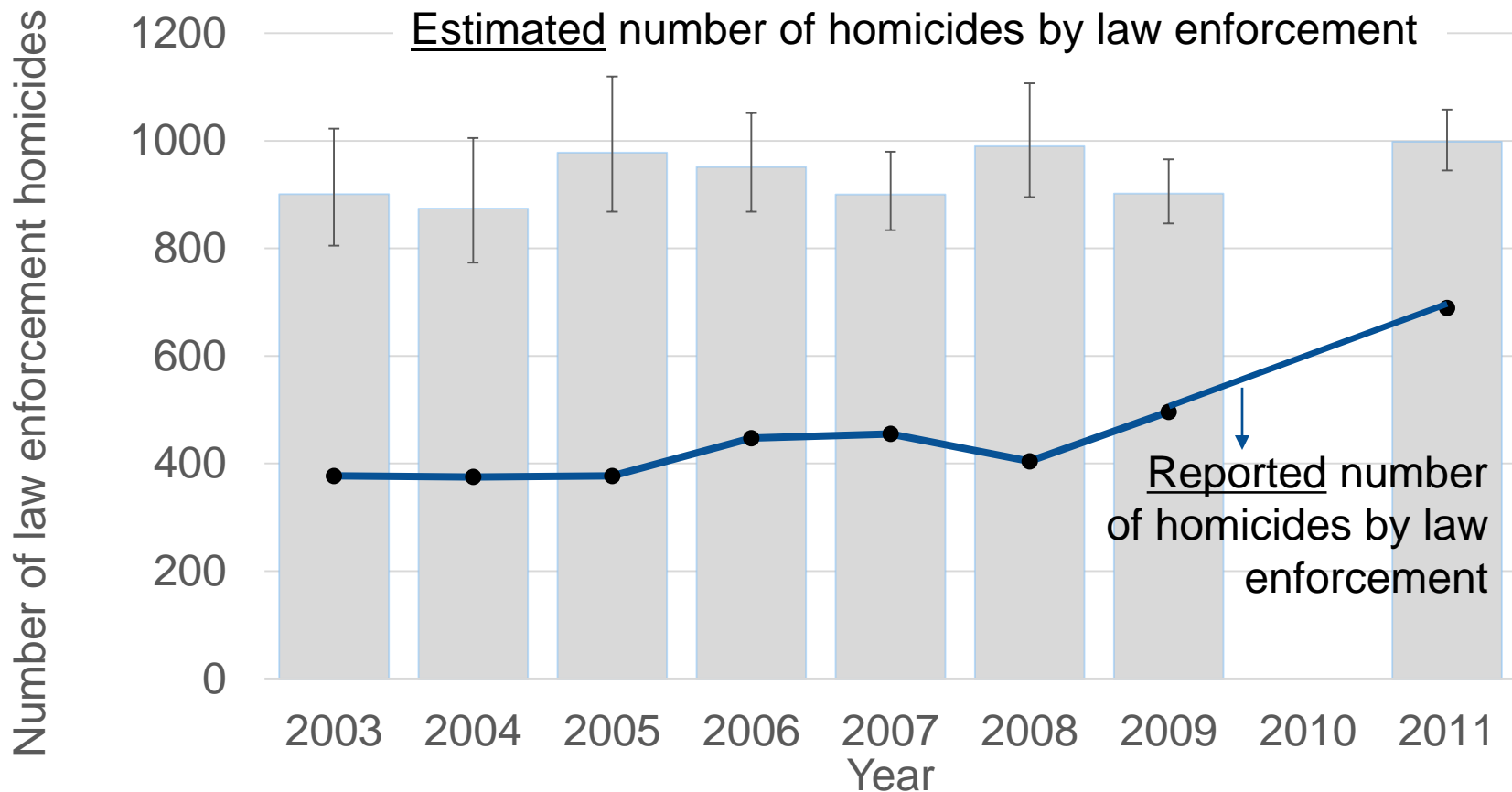
ARD Program Methodology, 2003-2011

BJS collected ARD program data from 2003 through 2011 primarily through a State Reporting Coordinator (SRC) model:

- A central reporter in each state was responsible for identifying and compiling information about each qualifying death occurring during the process of arrest.
- BJS provided the SRCs with guidance on methods to use to identify, qualify, and report data.
- SRCs had some degree of flexibility in compiling this information and took advantage of the relevant sources available to them in their states.
- In 2011, BJS began to conduct media searches to identify deaths in addition to relying on reports from SRCs.

BJS coverage assessment findings

Number of homicides by law enforcement reported to the ARD program and estimated coverage, 2003–2009 and 2011



Source: Banks, D., Couzens, G.L. & Planty, M.G. (2015, October). *Assessment of Coverage in the Arrest-Related Deaths Program* (NCJ 249099). Bureau of Justice Statistics, US Department of Justice.

Hybrid Approach to Data Collection

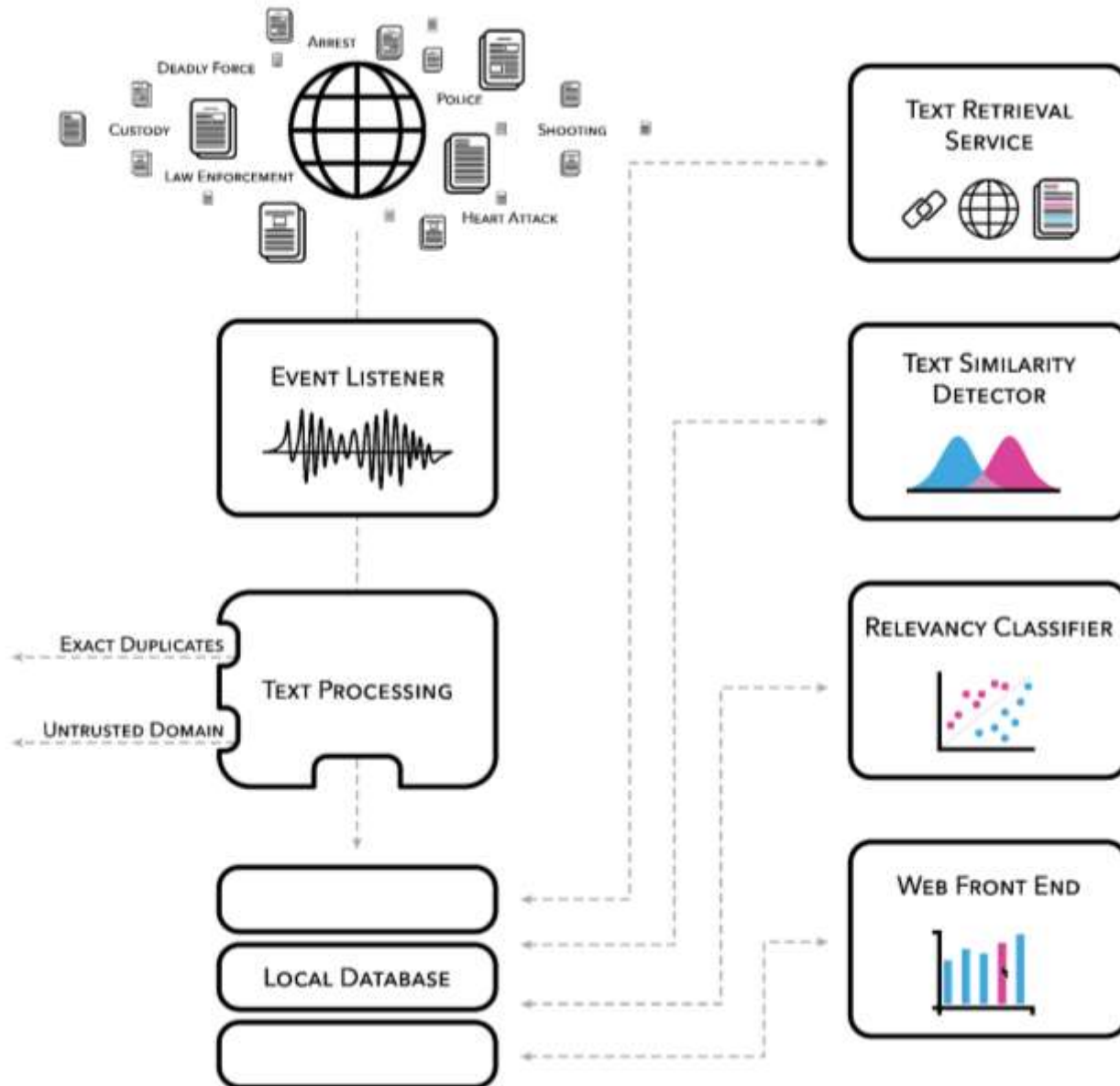
Phase 1: Identification

- Review news alerts and other public sources for potential arrest-related deaths.
- Track the efficiency and coverage of various approaches for identifying arrest-related deaths.
- Compile quarterly lists of potential arrest-related deaths.

Phase 2: Agency survey

- Conduct follow up with law enforcement agencies and medical examiner/coroner's offices to confirm identified deaths and collect more information about the circumstances surrounding the deaths.
- Survey agencies with identified deaths and others to assess and adjust Phase 1 identification methods.

Media monitoring, coding and classification pipeline



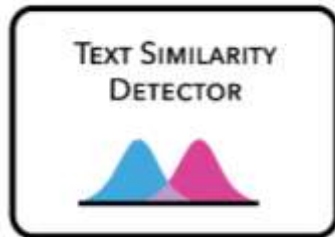
Event Listener



Returns articles with a combination of the primary search term and another term indicating law enforcement involvement such as "police," "officer," "arrested."

- Shoot(ing,er)
- Kill(ed,ing)
- Death, dead, died
- Deadly/lethal force
- Use of force
- Heart attack
- Accidental
- Overdose
- Taser, stun gun
- Standoff

Text Similarity Detector: TF-IDF Similarity



- Compute TF-IDF* similarity for all pairs of articles in 10-day sliding windows
- If a pair has high similarity, drop one of them
- Compare both texts and titles (above length limit)
- 270k/month reduction

*Term Frequency-Inverse Document Frequency

A former **South Carolina** police chief has dodged prison time in the 2011 shooting death of an unarmed Black man, receiving just one year of home detention, according to [NPR](#).

Former Eutawville, S.C. police chief **Richard Combs**, 38, shot and killed 54-year-old **Bernard Bailey** as he left the Eutawville Police Department in May 2011.

A former South Carolina police chief who was charged with murder in the shooting death of a black man in 2011 pleaded guilty to misconduct charges after two separate trials ended in hung juries.

The former Eutawville chief, Richard Combs, had been indicted more than three years after he shot and killed Bernard Bailey, 53, outside the police department. Combs argued that he feared Bailey would use a vehicle as a weapon against him.

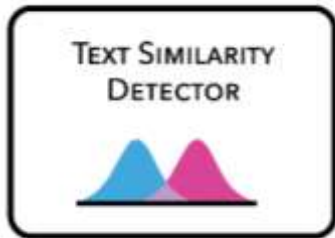
Relevancy Classifier



- Weighted combination of ensemble model (logistic regression and decision tree classifiers) and a neural network
- Separate models applied to text and title
- Words positively associated with arrest-related death
 - “shot and killed,” “armed,” “died,” “suspect”
- Words negatively associated with arrest-related death
 - “victim,” “anyone with,” “police investigate”

Remaining volume after all steps: 8,750 articles/month (99.2% reduction)

Text Similarity Detector: Entity Similarity



San Bernardino, California (CNN)—With the investigation still unfolding, much is unclear about Wednesday's deadly San Bernardino shooting at a center for people with developmental disabilities.

[See latest developments](#)

[Update 12:22 a.m ET Thursday]

The family has not been able to track down recently named suspect [Syed Farook](#) or his wife since Wednesday morning, said Hussam Ayloush, executive director of CAIR (the Council on American Islamic Relations).

- Generate a set of named entities mentioned in each article
- If the sets are similar enough, drop one article
- Compare texts only
- 105k/month reduction

Assailants Syed Farook and Tashfeen Malik killed in shootout with police after deadly attack at holiday party that killed 14 and injured at least 17 more

In the latest burst of deadly gun violence in the U.S., two heavily armed assailants entered a social-services agency in San Bernardino, California, on Wednesday and opened fire during a holiday party.

At least 14 people were killed in the initial attack and another 17 were injured, according to San Bernardino police officials.

About four hours later, local police located a dark-colored SUV and engaged in a shootout that killed two suspects, one male and one female. They were later identified as Syed Farook, 28, and Tashfeen Malik, 27, the latter believed to be Farook's girlfriend or wife.

Manual Article Review Portal

ARD

Last Decedent Reviewed: 183565 | Articles Reviewed Today: 3

Duren  Log Out

Agent: Officer had conflicting accounts of black man's death

MONTGOMERY, Ala. — A white Alabama police officer charged with killing a black man gave two conflicting accounts of what happened to a state investigator, the agent testified at a hearing Thursday. Officer Aaron Smith is charged with murder in the ...

 [Open source article in a new window](#)

- Does Not Meet ARD Definition
- Unknown Whether Meets ARD Definition (Needs Follow-Up)
- Meets ARD Definition
- Meets ARD Definition But Does Not Fall Within Our Date Scope

 Save and close

Follow-Up Notes

Notes

First Name

First name

Middle Name

Middle name

Last Name

Last name

Agency Name

Agency name

 [News](#) [Local and State](#) [Education](#) [Business](#) [Arts](#) [Health](#) [Entertainment](#) [Lifestyle Leisure](#)

24 March 2016 Thursday 17:49


This news clip was read 29 times.

Agent: Officer had conflicting accounts of black man's death

State Bureau of Investigation Agent Jason DiNunzio testified Smith didn't suspect Gunn of a crime when he initially stopped him, but Smith immediately confronted Gunn and told him to put his hands on the hood of the patrol car. DiNunzio said he doesn't...

 Share on Facebook  Tweet  Share on Google+

T- T+ A 

 Agent: Officer had conflicting accounts of black man's death

MONTGOMERY, Ala. — A white Alabama police officer charged with killing a black man gave two conflicting accounts of what happened to a state investigator, the agent testified at a hearing Thursday.

Officer Aaron Smith is charged with murder in the Feb. 25 death of 58-year-old Greg Gunn. After Thursday's hearing, Judge Jimmv Pool ruled there was probable cause

★ Feature



Matt Had
hustles t
Funny C

Gregory Gunn | Montgomery, AL | Montgomery PD | 2016-02-25 | (96%)

Select Known Decedent

Overall reduction in article volume

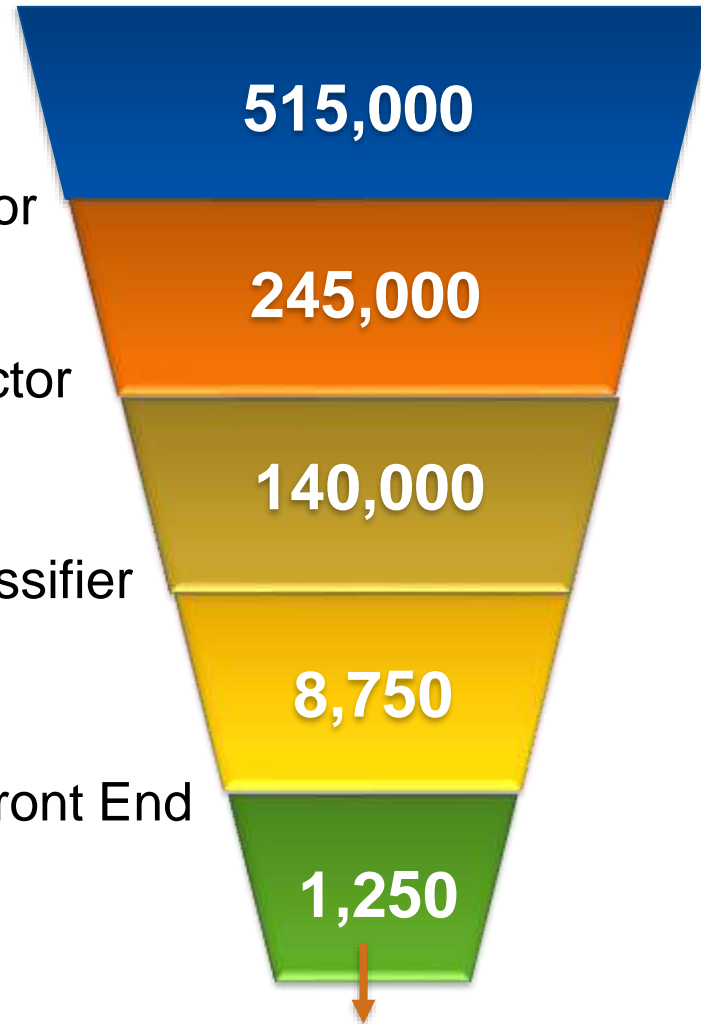
Typical monthly media alert volume
(after initial text processing)

TF-IDF Similarity Detector

Entity Similarity Detector

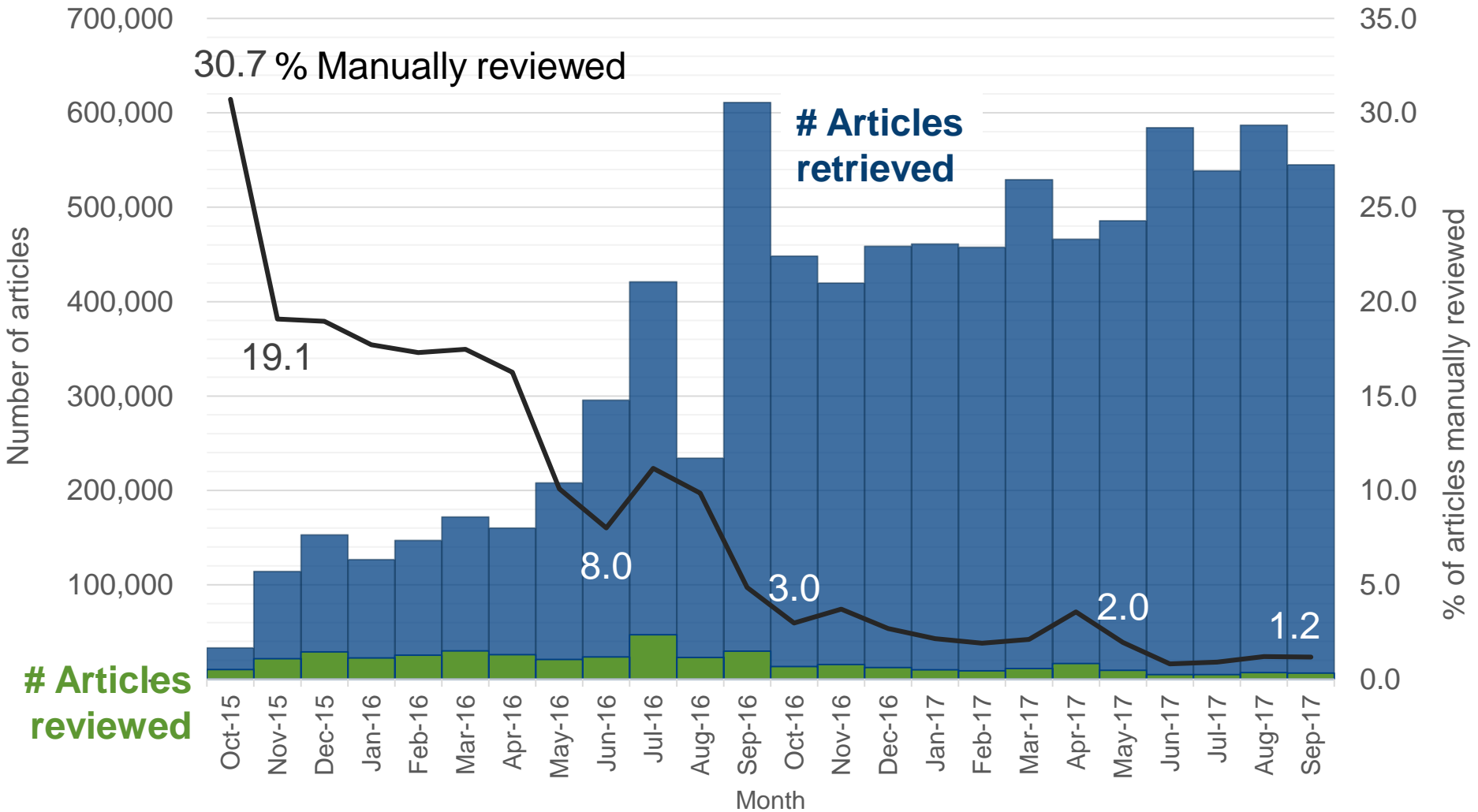
Relevancy Classifier

Web Front End

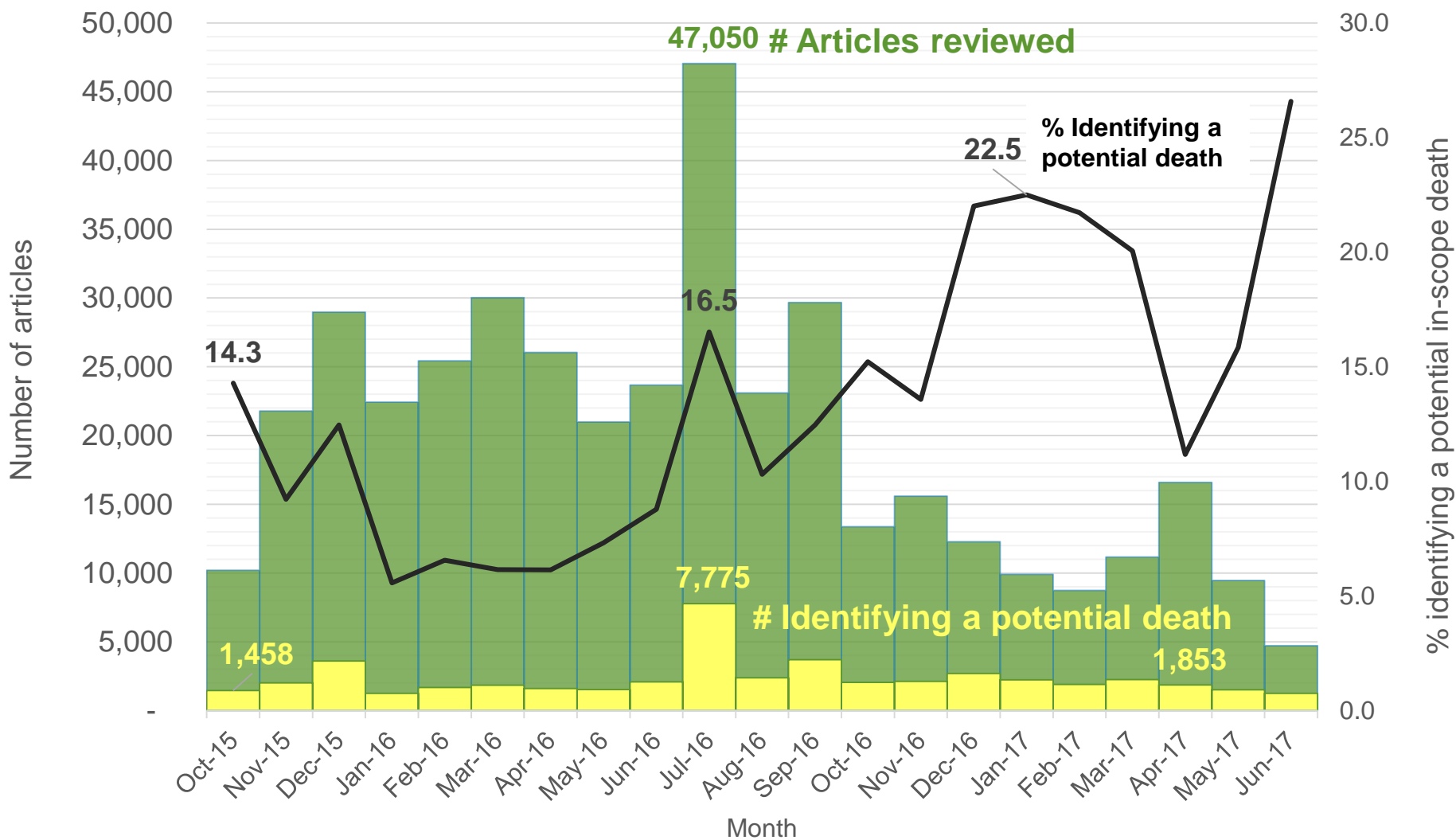


Average number of articles/month
identifying a potential arrest-related death

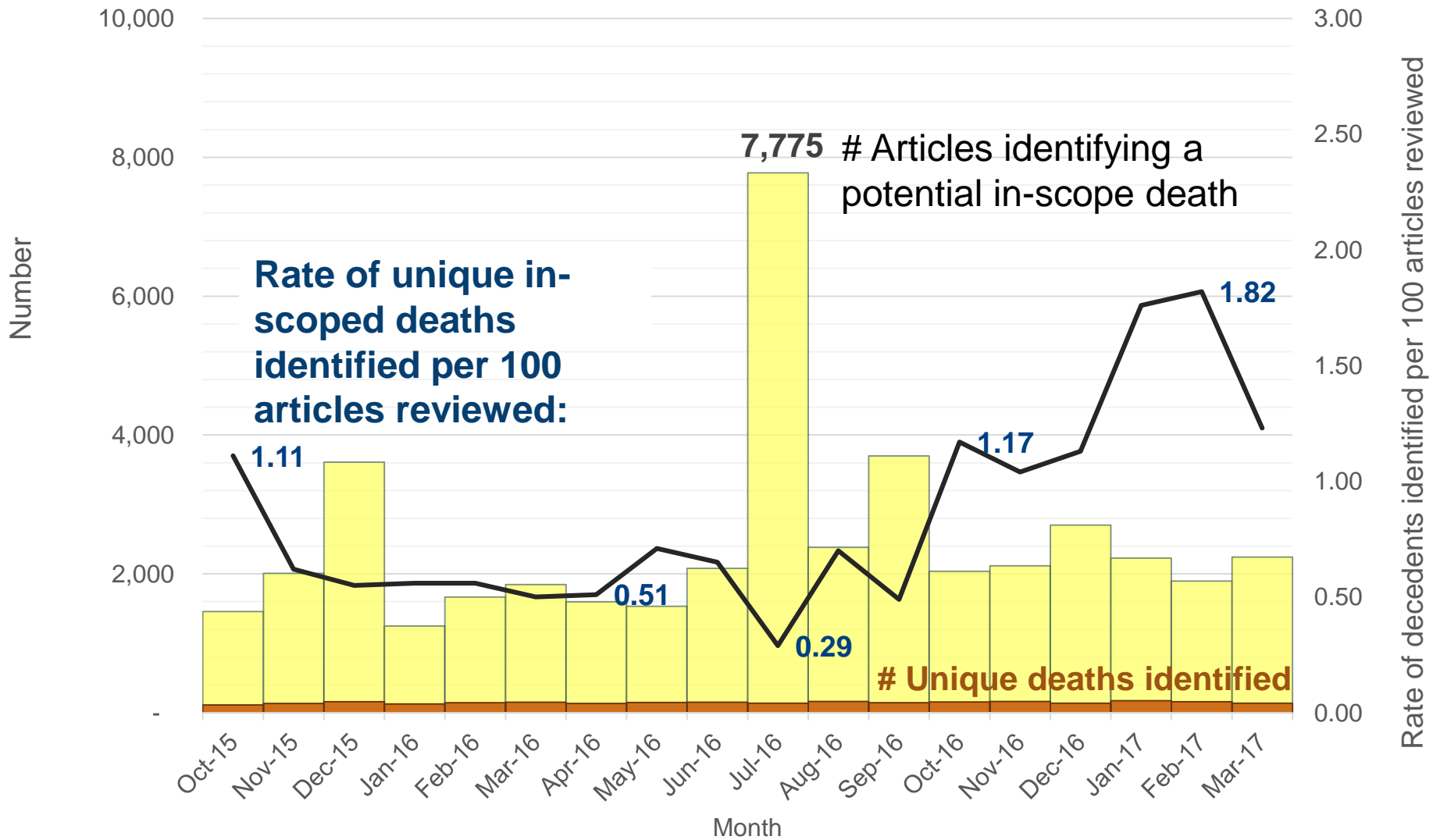
Number of articles retrieved, manually reviewed, and % of articles manually reviewed by month, Oct 2015 – Sep 2017



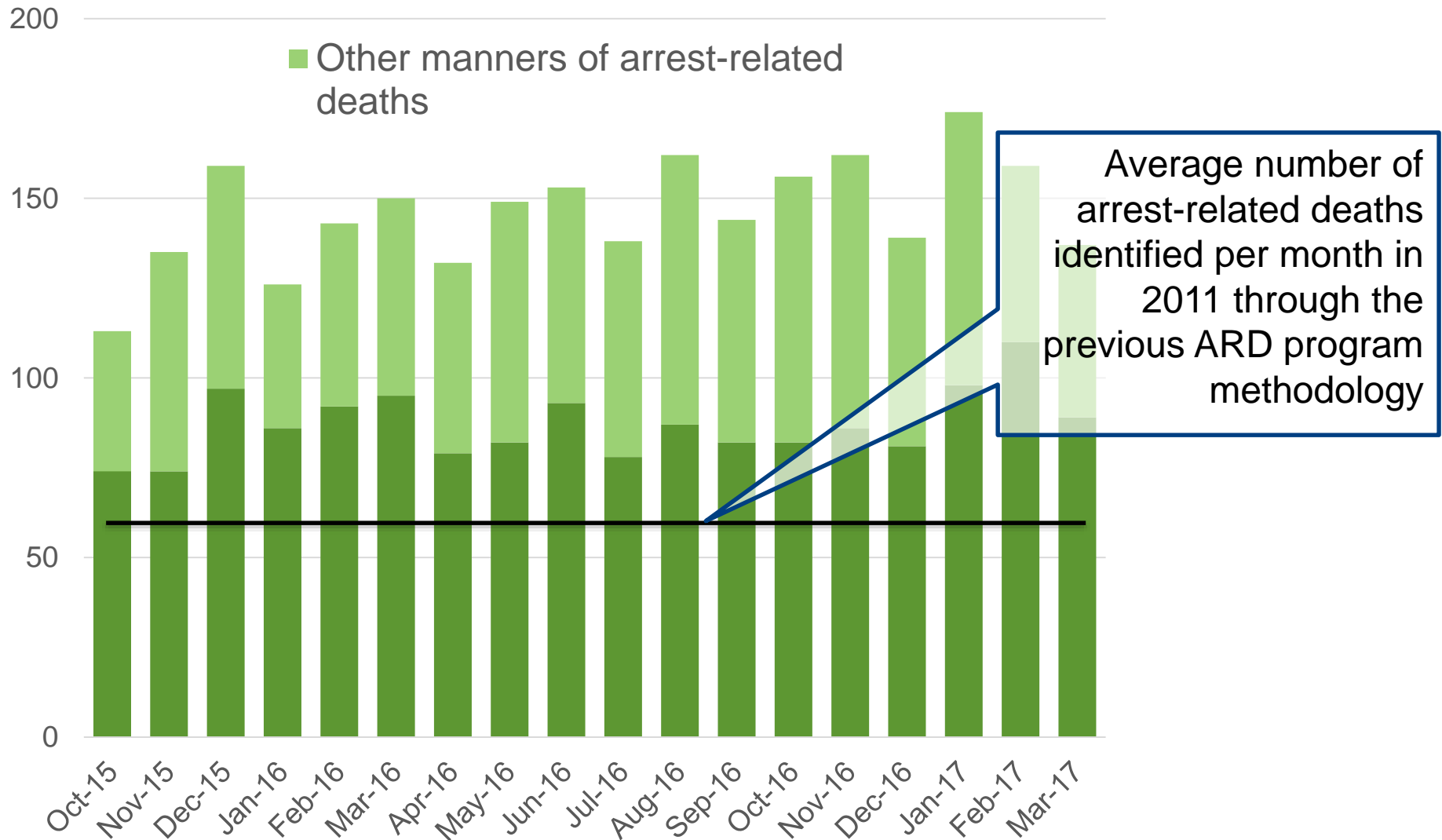
Number of articles manually reviewed, number and % identifying a potential in-scope death, Oct 2015 – Jun 2017



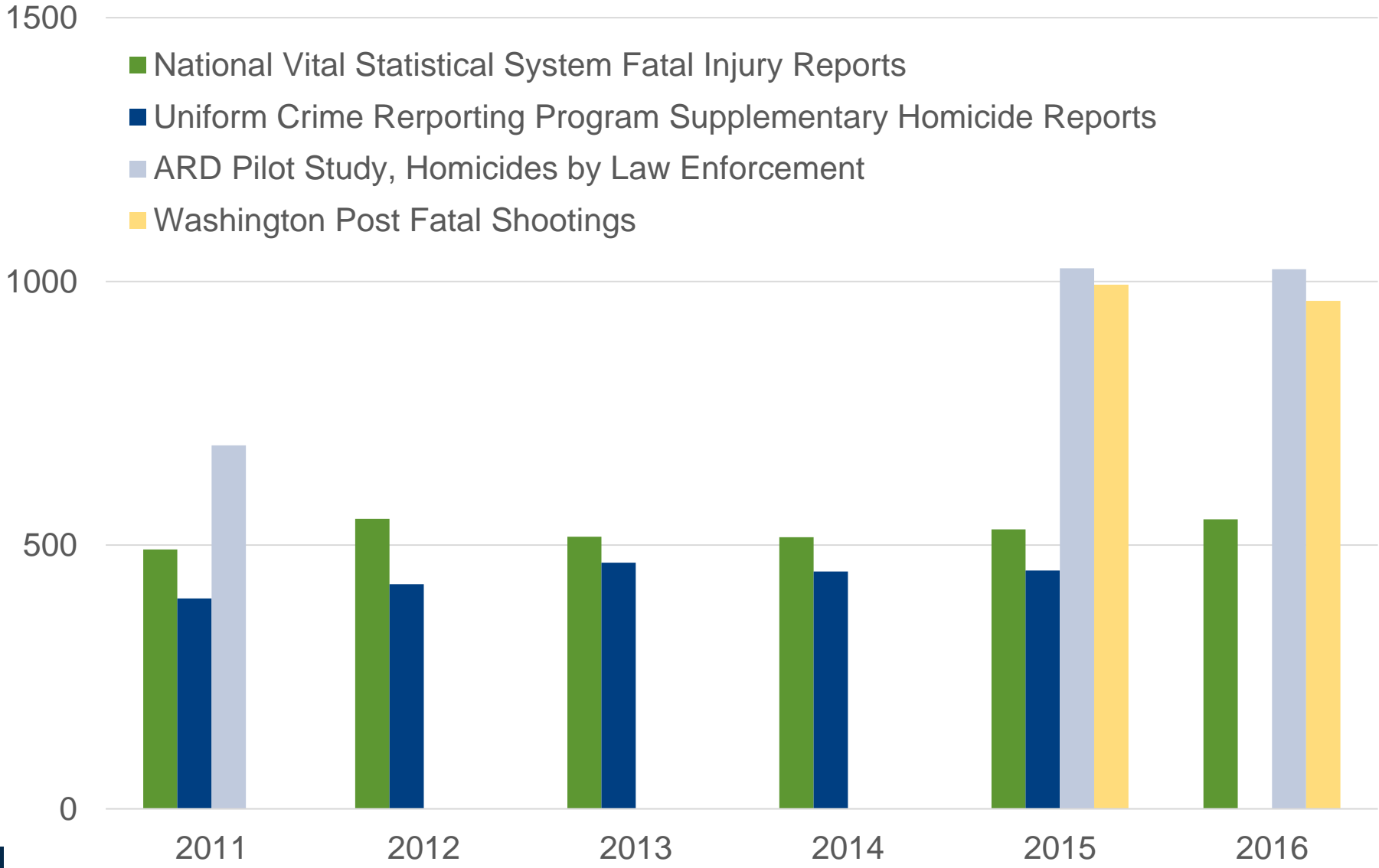
Number of articles identifying a potential in-scope death, number of unique deaths identified, and rate of unique deaths identified, Oct 2015 – Mar 2017



Unique decedents identified through open source by month and manner of death, October 2015 through March 2017



Number of homicides by law enforcement identified by year and source



Other considerations

- Goal: 100% coverage of all arrest-related deaths
- Monitoring of manual coding (quality control)
 - Further research for articles marked as “unknown” whether death meets ARD pilot study scope
 - “Known decedent” identification and linking
 - Resolving disagreements between coders and the Relevancy Classifier
 - Comparison to external sources
- Measuring deaths not covered in media sources

Hybrid Approach to Data Collection

Phase 1: Identification

- Review news alerts and other public sources for potential arrest-related deaths.
- Track the efficiency and coverage of various approaches for identifying arrest-related deaths.
- Compile a quarterly list of potential arrest-related deaths.

Phase 2: Agency survey

- Conduct follow up with law enforcement agencies and medical examiner/coroner's offices to confirm identified deaths and collect more information about the circumstances surrounding the deaths.
- Survey agencies with identified deaths and others to assess and adjust Phase 1 identification methods.

Law Enforcement Agencies and Medical Examiner/ Coroner's Offices Were Asked to

- Confirm all case-identifying data, including decedent name, decedent date of death, and name of LEA with jurisdiction over the decedent at the time of death *for decedents identified from media sources*
- Identify any arrest-related deaths occurring during study period that were not identified through the media sources
- Provide information about the decedent and factors precipitating law enforcement involvement with decedent for all confirmed arrest-related deaths.

ARD Pilot Study Incident Form Measures

Form CJ-11A

Law Enforcement Agency Arrest-related Death Incident Report

- Identifying information
- Location of incident
- Decedent demographics
- Precipitating events
- Decedent behavior during incident
- Law enforcement actions during incident
- Number of officers and LEAs that responded to incident

Form CJ-12A

Medical Examiner/Coroner's Office Arrest-related Death Incident Report

- Identifying information
- Location of incident
- Decedent demographics
- Whether autopsy was performed
- Manner of death
- Cause of death
- If died from injuries, how were those injuries sustained?
- If weapon caused death, what type of weapon?

Discussion

- Methodology for reliable and valid prevalence data on law enforcement homicides and other arrest-related deaths over time
- What else might be suitable to measure by open source review?
 - Police use of force
 - Hate/bias crimes
 - School shootings
 - Mass shooting incidents
 - Terrorist acts
- Open source vs. crowd sourced
- Limits of machine learning/open source methods