

Analytic Fusion for Essential Indicators of the Opioid Epidemic

Howard Burkom¹, Joseph Downs¹, Raghav Ramachandran¹, Wayne Loschen¹, Laurel Boyd², Matthew Laidler², Joshua Van Otterloo²

¹Research and Exploratory Development, Johns Hopkins Applied Physics Laboratory, Baltimore, Maryland, United States, ²Oregon Health Authority, Public Health Division, Salem, Oregon, United States

Objective

In a partnership between the Public Health Division of the Oregon Health Authority (OHA) and the Johns Hopkins Applied Physics Laboratory (APL), our objective was develop an analytic fusion tool using streaming data and report-based evidence to improve the targeting and timing of evidence-based interventions in the ongoing opioid overdose epidemic. The tool is intended to enable practical situational awareness in the ESSENCE biosurveillance system to target response programs at the county and state levels. Threats to be monitored include emerging events and gradual trends of overdoses in three categories: all prescription and illicit opioids, heroin, and especially high-mortality synthetic drugs such as fentanyl and its analogues. Traditional sources included emergency department (ED) visits and emergency management services (EMS) call records. Novel sources included poison center calls, death records, and report-based information such as bad batch warnings on social media. Using available data and requirements analyses thus far, we applied and compared Bayesian networks, decision trees, and other machine learning approaches to derive robust tools to reveal emerging overdose threats and identify at-risk subpopulations.

Introduction

Unlike other health threats of recent concern for which widespread mortality was hypothetical, the high fatality burden of opioid overdose crisis is present, steadily growing, and affecting young and old, rural and urban, military and civilian subpopulations. While the background of many public health monitors is mainly infectious disease surveillance, these epidemiologists seek to collaborate with behavioral health and injury prevention programs and with law enforcement and emergency medical services to combat the opioid crisis. Recent efforts have produced key terms and phrases in available data sources and numerous user-friendly dashboards allowing inspection of hundreds of plots. The current effort seeks to distill and present combined fusion alerts of greatest concern from numerous stratified data outputs. Near-term plans are to implement best-performing fusion methods as an ESSENCE module for the benefit of OHA staff and other user groups.

Methods

By analyzing historical OHA data, we formed features to monitor in each data source to adapt diagnosis codes and text strings suggested by CDC's injury prevention division, published EMS criteria [Reference 1], and generic product codes from CDC toxicologists, with guidance from OHA Emergency Services Director David Lehrfeld and from Oregon Poison Center Director Sandy Giffen. These features included general and specific opioid abuse indicators such as daily counts of records labelled with the "poisoning" subcategory and containing "fentanyl" or other keywords in the free-text. Matrices of corresponding time series were formed for each of 36 counties and the entire state as inputs to region-specific fusion algorithms.

To obtain truth data for detection, the OHA staff provided guidance and design help to generate plausible overdose threat scenarios that were quantified as realistic data distributions of monitored features accounting for time delays and historical distributions of counts in each data source. We sampled these distributions to create 1000 target sets for detection based on the event duration and affected counties for each event scenario.

We used these target datasets to compare the detection performance of fusion detection algorithms. Tested algorithms included Bayesian Networks formed with the R package gRain, and also random forest, logistic regression, and support vector machine models implemented with the Python scikit-learn package using default settings. The first 800 days of the data were used for model training, and the last 400 days for testing. Model results were evaluated with the metrics:

Sensitivity = (number of target event days signaled) / (all event days) and Positive predictive value (PPV) = (number of target event days signaled) / (all days signaled).



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

These metrics were combined with specificity regarded as the expected fusion alert rate calculated from the historical dataset with no simulated cases injected.

Results

The left half of Figure 1 illustrates a threat scenario along Oregon’s I5 corridor in which string of fentanyl overdoses with a few fatalities affects the monitored data streams in three counties over a seven-day period. The right half of the figure charts the performance metrics for random forest and Bayesian network machine learning methods applied to both training and test datasets assuming total case counts of 50, 20, and 10 overdoses. Sensitivity values were encouraging, especially for the Bayesian networks and even for the 10-case scenario. Computed PPV levels suggested a manageable public health investigation burden.

Conclusions

The detection results were promising for a threat scenario of particular concern to OHA based on a data scenario deemed plausible and realistic based on historical data. Trust and acceptance from public health surveillance of outputs from supervised machine learning methods beyond traditional statistical methods will require user experience and similar evaluation with additional threat scenarios and authentic event data.

Credible truth data can be generated for testing and evaluation of analytic fusion methods with the advantages of several years of historical data from multiple sources and the expertise of experienced monitors. The collaborative generation process may be standardized and extended to other threat types and data environments.

Next steps include the addition to the analytic fusion capability of report-based data that can influence data interpretation, including mainstream and social media reports, events in neighboring regions, and law enforcement data.

Acknowledgement

We would like to acknowledge the Hopkins APL Internal Research and Development program and substantial support from the following programs within the Oregon Public Health Division: Emergency Medical Services and Trauma Systems Program, Center for Health Statistics, Acute and Communicable Disease Prevention Program, and Injury and Violence Prevention Program along with support from the Oregon Poison Center, the National Poison Data System and the Injury Prevention & Control Center at the Centers for Disease Prevention and Control.

References

1. Rhode Island Enhanced State Opioid Overdose Surveillance (ESOOS) Case Definition for Emergency Medical Services (EMS), <http://www.health.ri.gov/publications/guidelines/ESOOSCaseDefinitionForEMS.pdf>, last accessed: Sept. 9, 2018.

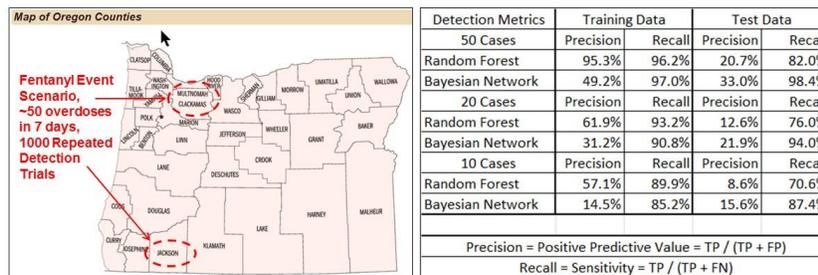


Figure 1: Schematic depicting overdose scenario randomized for Monte Carlo trials, with detection performance metrics tabulated for trials of 50, 20, and 10 cases over a 7-day period.



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.