Early Intervention Systems: Predicting Adverse Interactions Between Police and the Public

Jennifer Helsby[1], Samuel Carton[2], Kenneth Joseph[3], Ayesha Mahmud[4], Youngsoo Park[5], Andrea Navarrete[1], Klaus Ackermann[1], Joe Walsh[1], Lauren Haynes[1], Crystal Cody[6], Major Estella Patterson[6], and Rayid Ghani[1]

Abstract

Adverse interactions between police and the public hurt police legitimacy, cause harm to both officers and the public, and result in costly litigation. Early intervention systems (EISs) that flag officers considered most likely to be involved in one of these adverse events are an important tool for police supervision and for targeting interventions such as counseling or training. However, the EISs that exist are not data-driven and based on supervisor intuition. We have developed a data-driven EIS that uses a diverse set of data sources from the Charlotte-Mecklenburg Police Department and machine learning techniques to more accurately predict the officers who will have an adverse event. Our approach is able to significantly improve accuracy compared with their existing EIS: Preliminary results indicate a 20% reduction in false positives and a 75% increase in true positives.

[1]University of Chicago, IL, USA
[2]University of Michigan, Ann Arbor, USA
[3]Northeastern University, Boston, MA, USA
[4]Princeton University, NJ, USA
[5]The University of Arizona, Tucson, USA
[6]Charlotte-Mecklenburg Police Department, North Carolina, USA

Corresponding Author:
Jennifer Helsby, Computation Institute, University of Chicago, 5735 South Ellis Avenue, Chicago, IL 60637-5418, USA.
Email: jen@redshiftzero.com

Introduction

Recent high-profile cases of police officers using deadly force against members of the public have caused a political and public uproar (e.g., "Timeline," 2015; "Topic," 2016). They have also highlighted and increased tensions between the U.S. police force and citizens. While such violent altercations tend to capture the nation's attention, there is evidence that more mundane interactions between the police and the public can have negative implications as well (Jones, 2014).

Adverse events between the police and the public thus come in many forms, from deadly use of a weapon to a lack of courtesy paid to a victim's family. These events can have negative mental, physical, and emotional consequences on both police officers and citizens. We discuss our precise definition of "adverse event" below as an aspect of our experimental design.

Prior work has shown that a variety of factors predict adverse events (Arthur, 2015; Goldstein, 1977). While some of these factors are beyond the control of police officers and their departments, many of them can theoretically be addressed ahead of time. For example, training in appropriate use of force may reduce the odds of an officer deploying an unnecessary level of force in a particular situation.

The incidence of such factors is neither randomly distributed among officers nor over time (Goldstein, 1977). Certain officers, at certain periods of time, can be identified as being more at risk of involvement in an adverse event than others. Because police departments have limited resources available for interventions, a system to identify these high-risk officers is vital. Using this kind of Early Intervention System (EIS), police departments can provide targeted interventions to prevent adverse events, rather than being reactive and dealing with them after such an event occurs.

The work described in this article was initiated as part of the White House's Police Data Initiative, launched based on President Obama's Task Force on 21st-Century Policing. As part of this effort, we discussed EISs with several U.S. police departments, and it became clear that existing EISs were ineffective at identifying at-risk officers. This article describes our work with the Charlotte-Mecklenburg Police Department (CMPD) in North Carolina to use machine-learning algorithms—computer algorithms that learn patterns from historical data—to improve their existing EIS.

CMPD's 1,800 officers patrol more than 500 sq mi encompassing more than 900,000 people. Over the last 10 years, CMPD has become a leader in data-driven policing by investing heavily in a centralized data warehouse and building its own software, including an EIS. Like most EISs, CMPD's system uses behavioral thresholds, chosen through expert intuition, to flag officers. The officer's supervisor then determines whether an intervention is appropriate. Several departments have adopted CMPD's system as it was built more than 10 years ago (Shultz, 2015). To improve the current system, we focus on the following prediction task:

Given the set of all active officers at a given date (typically today) and all data collected by a police department prior to that date, predict which officers will have an adverse interaction in the next year.
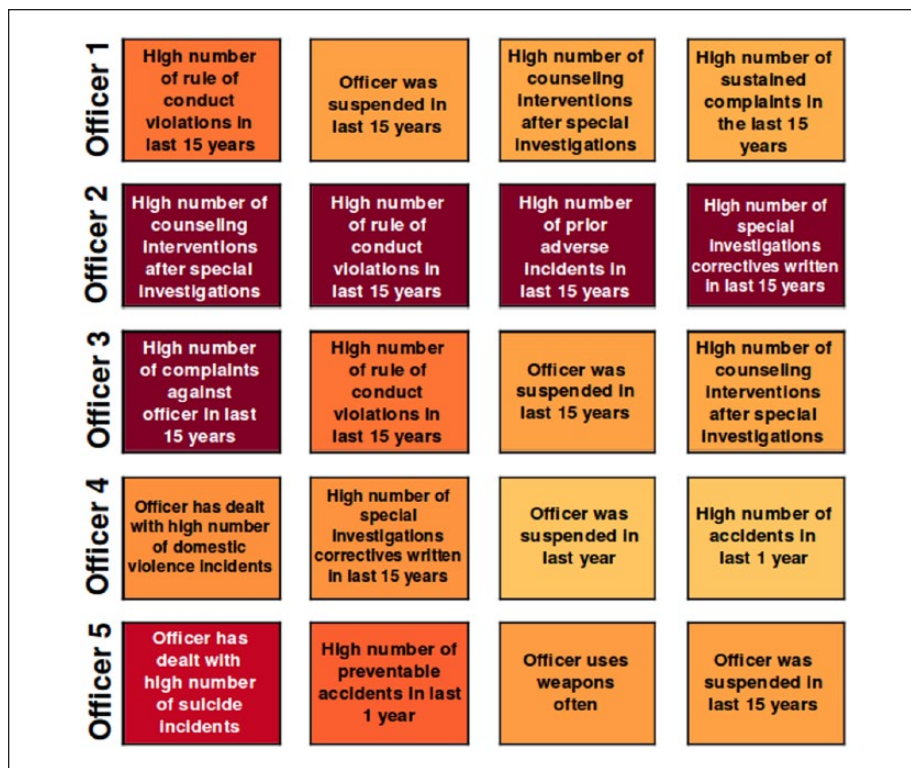
Figure 1. An illustration of five at-risk officers who will go on to have an adverse incident and their risk factors.
Note. The darker the red, the stronger the importance of that feature.

Although the work in this article is focused on predicting which officers will have an adverse interaction in the next year, we believe that our approach generalizes to different time horizons in the future. We show in this work that a machine learning model with an extensive set of indicators significantly outperform the department's existing EIS. Specifically, based on backtesting, our predictive model shows a relative increase of ~75% in true positive rate and a relative decrease of ~22% in false negative rate over the existing EIS. Unlike the existing system, our approach uses a data-driven approach and can thus be used to explore officer characteristics and neighborhood and environmental factors that predict adverse events beyond the handful of indicators used in existing EIS.

Figure 1 depicts five officers the EIS flagged as high risk, as well as their risk factors. Each officer in Figure 1 went on to have an adverse event. These risk factors were met with substantial acceptance by CMPD—an indicator of external validity of our modeling approach. The indicators vary for each officer, which highlights the need for

an extensive set of indicators to be provided to the system to have accurate predictions.

The system described here is the beginning of an effort that has the potential to allow police chiefs and supervisors across the nation to see which of their officers are in need of training, counseling, or additional assistance to make them better prepared to deal safely and positively with individuals and groups in their communities. Police departments can move from being responsive to negative officer incidents to being proactive and preventing these adverse incidents from happening in the first place.

In summary, the contributions of this article are the following:

- We apply, to our knowledge, the first adaptive, data-driven EIS that applies machine learning to predict adverse incidents from internal police department data.
- We show significant improvement over existing systems at flagging at risk officers.

Existing EISs

A small minority of officers account for the majority of adverse events, such as citizen complaints or excessive uses of force (Arthur, 2015; Goldstein, 1977). EISs, which are designed to detect officers exhibiting alarming behavioral patterns and prompt intervention such as counseling or training before serious problems arise, have been regarded as risk-management tools for countering this issue. The U.S. Commission on Civil Rights (1981), the Commission on Accreditation for Law Enforcement Agencies (2001), U.S. Department of Justice (2001), the International Association of Chiefs of Police, and the Police Foundation have recommended departments use EISs. Most federal consent decrees (legal settlements between the Department of Justice and a police department) issued to correct problematic policing require an EIS to be in place (Walker, 2003). A 2007 Law Enforcement Management and Administrative Statistics (LEMAS) survey showed that 65% of surveyed police departments with 250 or more officers had an EIS in place (Shjarback, 2015).

Existing EISs detect officers at risk of adverse events by observing a number of indicators and raising a flag when certain selection criteria are met. These criteria are usually thresholds on counts of certain kinds of incidents over a specified time frame, such as two accidents within 180 days or three uses of force within 90 days. Thresholds such as these fail to capture the complex nature of behavioral patterns and the context in which these events play out. For example, CMPD's system uses the same thresholds for officers working the midnight shift in a high-crime area and officers working in the business district in the morning.

More sophisticated systems flag outliers while accounting for one or two variables (as context), such as the officer's beat,[1] but still fail to include many factors. For example, CMPD's indicators include complaints, uses of force, vehicle pursuits and accidents, rule-of-conduct violations, raids and searches of civilians or civilian property, and officer injuries. Important factors, such as prior suspensions from the force, are often not included.

Empirical studies on the effectiveness of these systems have been limited, and their findings give mixed conclusions. Case studies focusing on specific police departments have shown that EISs were effective in decreasing the number of citizen complaints (Davis, Henderson, Mandelstam, Ortiz, & Miller, 2005; Walker, Alpert, & Kenney, 2001), but it is unclear whether this decrease arises from a reduction in problematic behavior or from discouraging officers from proactive policing (Worden et al., 2013). A large-scale study of emerging EISs across departments concludes that EIS effectiveness depends on departmental characteristics and details of implementation, such as which indicators are tracked, what thresholds are assigned, and how supervisors handle the system's flags (Shjarback, 2015). Beyond their possible ineffectiveness, threshold-based systems pose additional challenges. First, inconsistent use of the system creates an obstacle for threshold-based EISs. Second, threshold-based systems are difficult to customize. At least one vendor hard-codes thresholds into their EIS, making changes difficult and costly. Ideally, the system should improve as the department collects more data, but threshold-based systems require extensive use of heuristics, making such changes unlikely. Third, threshold systems are easily gamed. Because thresholds are visible and intuitive, officers can modify their behaviors slightly to avoid detection— either not taking an action they should have taken, or by not reporting an action they did take. Finally, output from threshold systems is limited to binary flags instead of risk scores. Risk scores enable the agency to rank people/facilities/and so forth by risk, to explicitly choose trade-offs—for example, increasing the number of false positives (officers who are flagged by the system but will not go on to have an adverse incident) to capture more true positives (officers who are flagged by the system and will go on to have an adverse incident), and to allocate resources in a prioritized manner.

A machine learning system can alleviate many of these issues. With respect to customization, machine learning models can be easily retrained on new data and with new variables. Furthermore, given the volume of variables and variable interactions that can be used within a machine learning model, parameters are sufficiently complex that the system cannot be easily gamed. Finally, machine learning approaches can be used to generate risk scores as opposed to pure binary classification. In addition to being a better fit for the resource constraints faced by today's U.S. police force, risk score systems can identify which officers are doing well as easily as which are at risk. The department can use this information when assigning officers to partners or when looking for best practices to incorporate into its training programs. When coupled with, for example, police-worn body camera footage, this system could be an important new tool for improving police practices.

Police Misconduct

Designing an effective EIS requires knowledge of what factors may predict adverse events. The literature on police behavior and misconduct has focused on three broad sets of variables: officer characteristics, situational factors, and neighborhood factors.

More educated police officers, particularly those with 4-year college degrees, tend to have fewer complaints and allegations of misconduct compared with officers with less education (Chapman, 2012; Manis, Archbold, & Hassell, 2008; White & Kane,

2013). In a study of misconduct in the New York Police Department, White and Kane (2013) found that, in addition to education level, prior records of criminal action, prior poor performance, and a history of citizen complaints were all significant predictors of misconduct as well.

Situational factors are those specific to particular incidents that (perhaps) result in an adverse event. These factors include demographics and behaviors of the citizen(s) involved in that particular incident as well as features` of the incident itself, such as time of day and location. White (2002) found that certain categories of incidents, such as robberies and disturbance incidents, were more likely to result in police use of deadly force. However, studies examining the relationship between citizen characteristics (such as race, gender, and age) and police behavior (such as likelihood of arrests and citations, and use of force) have found mixed results (Sobol et al., 2013). Research on citizen characteristics has, moreover, been limited due to lack of publicly available data.

Finally, neighborhood factors have been studied as a potential predictor of police misconduct. Sobol et al., (2013) found that incidents in high-crime neighborhoods have a greater likelihood of ending in interrogation, search, or arrest. Similarly, Terrill and Reisig (2003) found that police officers were more likely to use higher levels of force in disadvantaged and high-crime neighborhoods.

Our models incorporate variables for each of these levels of analysis, finding that variables at each level have a unique and important role in predicting officers at risk of adverse events. We are currently involved in efforts to experimentally distinguish causal factors. In the present work, however, efforts are restricted to understanding only those factors correlated with officers at risk of adverse events.

Data Description

The data for this work consist of almost all employee information and event records collected by CMPD to manage its day-to-day operations. The data contain records for all department employees since at least 2002, including those who retire, leave, or are terminated, but certain information, such as employee names, identification (ID) numbers, and military veteran status, as well as all narrative fields in the data, were redacted in accordance with North Carolina personnel laws to protect employee privacy and safety. The major types of information present in the data set, summarized in Table 1, are described in detail in this section. Almost all records are associated with one or more involved officers and include a hashed version of the ID of that officer in addition to any other information.

Internal Affairs (IA) Data

IA records contain the information about adverse events that we use as our *outcome variable*—the variable we are trying to predict. Every IA record pertains to a single officer. When a department employee or member of the public files a complaint or

Table 1. Description of the Types of Data Used, as well as the Number of Records Over Which Time Period.

Database	Number of records	Time window
Training	1.4M	2001–now
Internal Affairs	20K	2002–now
Traffic stops	1.6M	2002–now
Employee records	20K	2002–now
Dispatch events	14M	2003–now
Field interviews	180K	2003–now
Criminal complaints	959K	2005–now
Existing EIS	14K	2005–now
Arrests	350K	2005–now
Citations	946K	2006–now
Secondary employment	651K	2009–now

Note. EIS = early intervention system.

when an officer uses force, engages in a vehicle pursuit, gets into a vehicle accident, commits a rule-of-conduct violation, is injured, or conducts a raid and search, CMPD creates an IA record. Each record contains additional information such as a link to the *dispatch event*[2] during which the incident took place. Finally, each record contains the reviewing supervisor's decision regarding the appropriateness of the officer's actions as well as the recommended intervention if intervention was deemed necessary.

IA investigations of different event types can carry different outcomes: Complaints can be deemed *sustained* or *not sustained*, accidents and injuries can be deemed *preventable* or *not preventable*, and everything else (e.g., use of force) can be deemed *justified* or *not justified*. We assign a 1 to the outcome variable if the officer has an *unjustified*, *preventable*, or *sustained* disposition within a year, except for a number of internal complaints that we consider less egregious, such as misuse of sick leave. Figure 2 shows the IA process and our definition for an adverse incident, and Table 2 lists the full set of IA outcomes that we label as adverse events.

We proceed with the assumption that the IA data reasonably represents the true distribution of adverse events and officer fault. For various reasons, this assumption may be flawed. For example, many departments screen complaints before entering them into their IA system, and incidents have been reported in which officers do not faithfully record events. While CMPD encourages good data collection by punishing officers who fail to report adverse incidents, there is no guarantee of data accuracy. In addition, almost all IA cases are resolved internally without reference to an external agency. Unfortunately, without similarly comprehensive data from other police departments, it is difficult to estimate what effect these biases might have on the present work. We thus note this point as a condition on which the present analysis should be qualified and plan to investigate this further as we expand our work to other police departments.
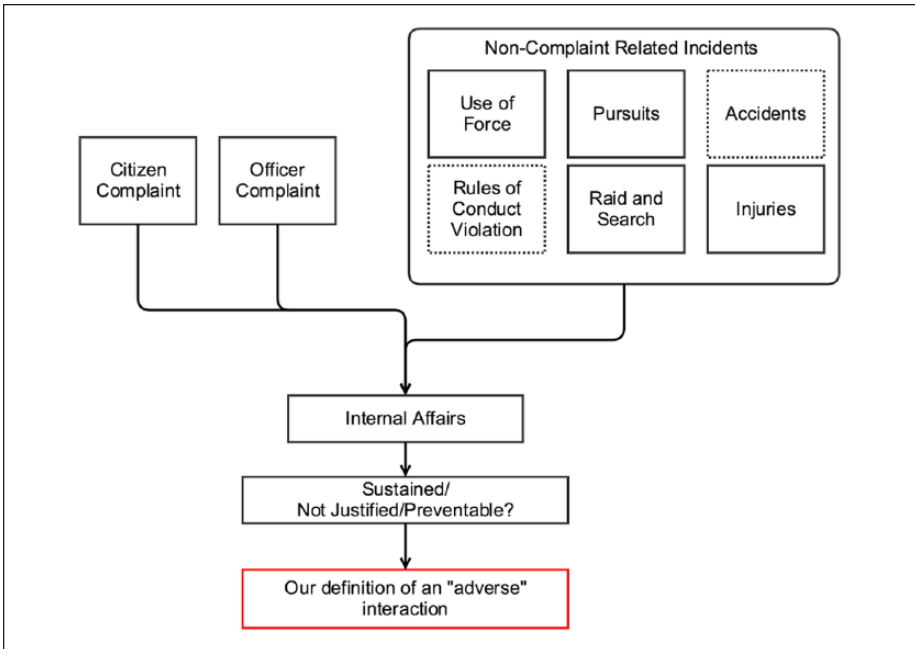
Figure 2. The IA process and our definition of an adverse incident.
Note. IA = internal affairs.

Table 2. The Types of Events Within the IA Database That We Define as Representative of an *Adverse Event*.

Event	IA ruling
Citizen complaint[a]	Sustained
Officer complaint[a]	Sustained
Vehicle accidents	Preventable
Injuries	Preventable
Use of force	Unjustified
Raid and search	Unjustified
Pursuit	Unjustified
Discharge of firearm	Unjustified
Tire deflation device	Unjustified

Note. IA = internal affairs.
[a]Minor violations excluded.

Other Data

Dispatch events. CMPD's system creates a dispatch event every time an officer is dispatched to a scene—for example, in response to a 911 call—and every time an

officer reports an action to the department. The dispatch system is the backbone of how officer movements are coordinated, and an officer's dispatches provide a rough guide to what the officer did and where the officer did it at all times they are active on the force. Dispatch records include the time and location of all events, as well as the type of event (e.g., robbery) and its priority. Dispatches are often linked in CMPD's system to other types of events, such as arrests or IA cases, that occurred during that dispatch.

Criminal complaints. The criminal complaints data provided by CMPD contains records of criminal complaints made by citizens. Each record includes a code for the incident, the location of the incident, the type of weapons involved if weapons were involved, and details about victims and responding officers. It also contains flags that include information such as whether the event was associated with gang violence, domestic violence, narcotics activity, or hate crimes.

Citations. The citations data provide details of each citation written by officers. Each record contains the date and type of citation, a code corresponding to the division, and additional metadata such as whether the citation was written on paper or electronically.

Traffic stops. CMPD officers are required to record information about all traffic stops they conduct. Records include time, location, the reason for and the outcome of the stop, if the traffic stop resulted in the use of force, and the stopped driver's sociodemographic profile.

Arrests. CMPD records every arrest made by its officers, including when and where the arrest took place, what charges were associated, whether a judge deemed the officer to have had probable cause, and the suspect's demographic information.

Field interviews. A "field interview" is the broad name given by CMPD for any event in which a pedestrian is stopped or frisked, or any time an officer enters or attempts to enter the property of an individual. In the latter case, officers may simply be completing a "knock and talk" to request information from a citizen, or be part of a team conducting a "raid and search" of an individual's property. A field interview can also be conducted as a result of a traffic stop. Records contain temporal and spatial information as well as information about the demographics about the interviewed person.

Employee records. The department's employee information includes demographic information such as officer education levels and years of service for every individual employed by the department, including those who have retired or been fired.

Secondary employment. CMPD records all events in which officers are hired by external contractors to provide security. These external contractors include, for example, financial institutions, private businesses, and professional sports teams. Officers are

allowed to sign up for these various opportunities through CMPD and are required to record all events that occur at them, such as disturbances, trespasses, or arrests.

Training. CMPD requires officers to receive rigorous training on a variety of topics, from physical fitness to how to interact with members of the public. The department records each officer's training events.

Existing EIS flags. We were also given the history of EIS flags going back over 10 years to 2005. Each record identifies the relevant officer and supervisor, the threshold triggered (e.g., more than two accidents in a 180-day period or more than three uses of force in a 90-day period), whether an action was taken in response to the flag, and if necessary, the selected intervention for each flag, which can include training and counseling.

Neighborhood. In addition to the data provided by CMPD, we also use publicly available data from 2010 and 2012 neighborhood quality-of-life studies[3] to understand the geospatial context of CMPD events. These studies collect data on many neighborhood variables, including Census/American Community Survey (ACS) data on neighborhood demographics and data on physical characteristics, crime, and economic vitality.

Data Limitations

In addition to the potential bias discussed above, the data set has a few other limitations. First, traffic stops, field interviews, and criminal complaints are entered into the CMPD system by the officers themselves, often in the midst of busy shifts or retroactively after their shifts have ended. Times and locations are often approximate, and these types of events often fail to be properly linked to an associated dispatch call, which limits what other information (such as IA cases) they can be linked to. Other important fields are also missing with relative frequency from the data. We take standard measures to accommodate missing data, and try to mitigate the unreliability of temporal and spatial information by aggregating the data across time and space in our predict generation.

Method

The goal of the EIS is to predict which officers are likely to have an adverse event in the near future. We use machine learning binary classification methods for this prediction task. Binary classification is a machine learning problem setting where we classify the entity of interest into one of two categories. In our case, we want to predict whether a given officer will have an adverse event in a given period of time into the future and formulate this as

- Outcome 0: An officer did not have an adverse incident in the next N days.
- Outcome 1: An officer had an adverse incident in the next N days.

Table 3. Categorization of Variables Used in the Data-Driven EIS.

Category	Variable count
Incident allegations	204
Traffic stops	200
Officer shifts	152
Officer arrests	136
Field interviews	120
Dispatches	116
Incident adjudication	108
Officer characteristics	32

Note. EIS = early intervention system.

In discussions with CMPD and in consideration of the rareness of adverse events, we decided that 1 year was an appropriate prediction window. In machine learning problems, a set of variables are defined, and an algorithm or *machine learning model* is used to determine how much each indicator should be weighted to best predict the probability that a given officer will have an outcome of 1.

Efforts were chiefly geared toward the extraction of these indicators from CMPD's data—in total 1,068 variables were used. Table 3 provides variable categorization. A plurality of variables relate to allegations about adverse incidents. The second category relates to traffic stops, the third to the officer's shift history (e.g., times and places), the fourth to the officer's arrests, the fifth to the officer's field interviews, and the sixth to the officer's dispatches. The next category includes information about the officer's adverse-incident adjudications. And the last includes characteristics about the officer, such as demographics and time on the force. For predictive modeling, we tried a variety of methods, including AdaBoost, Random Forests, Extra Trees, and Logistic Regression (for a full review of these models, please refer to Hastie, Tibshirani, & Friedman, 2009). Each machine learning model has a set of *hyperparameters* that tune the performance of the model. To select the best hyperparameters for the task, we search over the hyperparameter space to tune each model. Below, we discuss our process for generating variables and how the resulting models were evaluated.

Variable Generation

We generated variables based on our experience working on similar problems, as well as on discussions with experts at the CMPD. Patrol officers, IA investigators, members of our officer focus group, and department leadership suggested variables they believed may predict whether an officer has an adverse incident.

We generate behavioral variables by aggregating the record of incidents by each officer, establishing a behavioral history. The simplest variables are counts and

fixed-period counts of incidents the officer has been involved in (e.g., arrests, citations, etc.) and incident subtypes (e.g., arrests with only discretionary charges).

Notably among incident subtypes, we track incidents we believe are likely to contribute to officer stress, such as events involving suicides, domestic violence, young children, gang violence, or narcotics. In addition, we incorporate variables describing the number of credit hours of training officers had in topic areas of relevance: less-than-lethal weapons training, bias training, and physical fitness training.

To these counts, we add a variety of normalized and higher order variables. To account for high-crime times and locations, we include outlier variables, where we compare an officer's event frequencies against the mean frequencies for the officer's assigned division and beat. We generate time-series variables from raw event counts (e.g., a sudden increase in the number of arrests in the 6-month period prior to the point of analysis) to capture sudden changes in behavior. We also use more static officer variables such as demographics, height, weight, and time on the force.

Finally, we include neighborhood variables to capture specific information about the areas where officers patrol. For example, we included Charlotte-Mecklenburg's nonemergency services (311) call rate for CMPD patrol areas, which correlates not only with conditions in the neighborhood but also with the residents' willingness to report problems to city government.

Model Evaluation

We validate our models using temporal cross validation, or backtesting (Hyndman & Athanasopoulos, 2014). An example validation would be to train our models on data up to 2010 and then predict on the holdout set from all of 2011. As our data range from 2009 to 2016, we perform multiple evaluations over the data (train on 2009—predict on 2010, train on 2009 and 2010—predict on 2011, and so on) and aggregate them to come up with the final validation statistics.

For each evaluation, we use positive predictive value (percentage of officers flagged by our system who actually go on to have an adverse event in the next year) and sensitivity (percentage of officers with adverse events in the next year who are flagged by our system) at various population proportions as outcome metrics. For example, we can measure positive predictive value and sensitivity when we flag the top 5% of officers, as ordered by risk score.

Results

In this section, we discuss results in terms of predictive performance as well as an analysis of highly predictive variables. The best binary classification model to predict these events was an Extra Trees (Geurts et al., 2006) model. Extra Trees is an ensemble of decision trees learned from bootstrapped data, like a random forest, except the algorithm randomizes cutpoints in the constituent trees. The model includes 10,000 trees to reduce variance in which officers are flagged.[4]
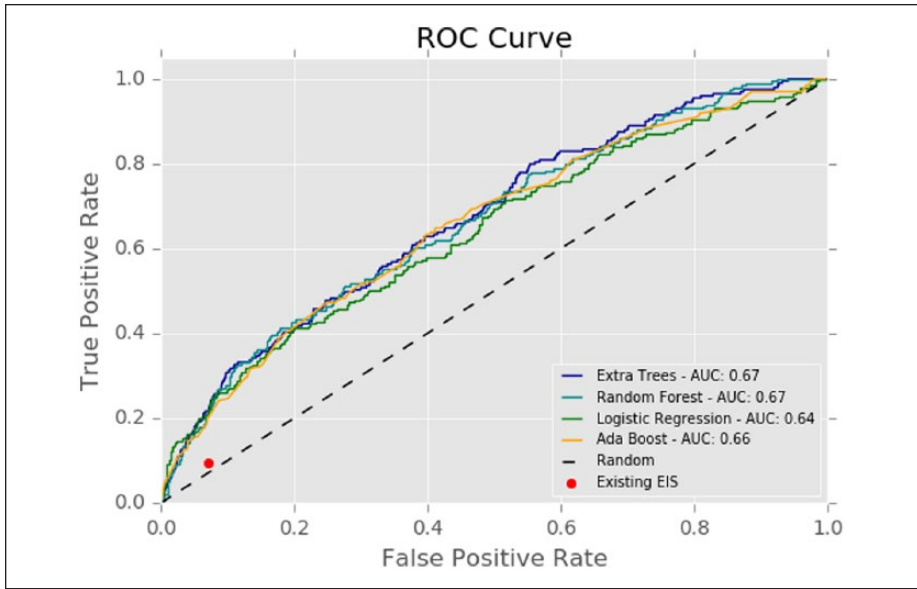
Figure 3. ROC curves for several machine learning models and the existing EIS.
Note. ROC = receiver operating characteristic; AUC = area under curve; EIS = early intervention system.

Predictive Performance

Figure 3 shows receiver operating characteristic (ROC) curves for several machine learning models, as well as the true positive rate and false positive rate for the existing EIS. All the machine learning algorithms presented significantly outperform CMPD's system, not only in accuracy but also flexibility: Given its use of clear-cut thresholds, the existing EIS makes choosing between true positives and false positives difficult. The existing EIS barely outperforms a random baseline. Extra trees, random forest, and AdaBoost yield higher areas under the ROC curve than logistic regression, which suggests complex underlying correlations. We use an extra trees model.

CMPD can choose how many officers it would like to flag. At one extreme, the department could choose to flag no officers. No officers would be incorrectly flagged, but at the same time no officers would be correctly flagged. The department would avoid wasting its limited resources on officers who do not need additional assistance, but it would also miss the opportunity to help prevent adverse events. At the other extreme, the department could flag all officers. All officers who go on to have adverse incidents would get flagged, but so would all the officers who do not go on to have adverse incidents. A high rate of false positives reduces the legitimacy of the system and prevents the EIS from helping the department distinguish which officers need assistance. Many EISs have been used inconsistently because they flag many officers who do not go on to have adverse incidents. The department needs to choose a

Table 4. Comparison of Model Performance Between the Existing Threshold-Based EIS and the Data-Driven EIS Between April 2, 2014 and April 1, 2015.

Metric	Existing EIS (flags 154)	Data-driven EIS (flag top 5%)	Data-driven EIS (flags 154)
True positives	34	31 (−9%)	60 (+76%)
False positives	120	74 (−38%)	94 (−22%)
True negatives	1,626	1,672 (+3%)	1,652 (+2%)
False negatives	323	326 (−1%)	293 (−9%)

Note. EIS = early intervention system.

threshold that does not violate its resource constraints and that gives tolerable true-positive and false-positive rates.

CMPD will modify its EIS procedures to monitor the performance of the EIS and to reduce unnecessary administrative burdens for supervisors and officers. Initially, EIS flags will go to a supervisor at headquarters, who will conduct a preliminary review before deciding whether to forward the flag to the officer's direct supervisor. With all flags going through a single supervisor, CMPD has decided to start by flagging 5% of officers a year but may adjust that number.

Table 4 shows how our model compares with the existing EIS in terms of false positives, false negatives, true positives, and true negatives when flagging (a) the top 5% of officers (CMPD's beginning strategy) and (b) the same number of officers as the existing EIS (154). Our results show that moving beyond the current threshold system and using a broader set of data with more complex models improves prediction performance. Our best performing model can flag 76% more high-risk officers (true positives), while flagging 22% fewer low-risk officers (false positives) compared with the current system.

Analysis of Important Indicators

Figure 4 shows the most important indicators found by our best performing model. The most predictive variables, not surprisingly, were those relevant to the prior IA history of the officer: Officers who are routinely found to have been engaged in an adverse event are likely to engage in another such event in the future. Even being accused of having an adverse event seems to predict adverse events. This is fairly typical in behavioral prediction tasks.

In addition, characteristics about the officer's type of work appear. Conducting field interviews where the citizen has a weapon correlate with future adverse incidents, as do traffic stops: traffic stops for safety violations, traffic stops for moving violations, traffic stops that result in a verbal warning, and traffic stops in general. We do not know the direction of these correlations—but it is likely conditional—nor do we claim a causal link.

Combined, these observations provide support for the idea that a subset of officers are at particular risk for adverse events, and that an EIS that controls for
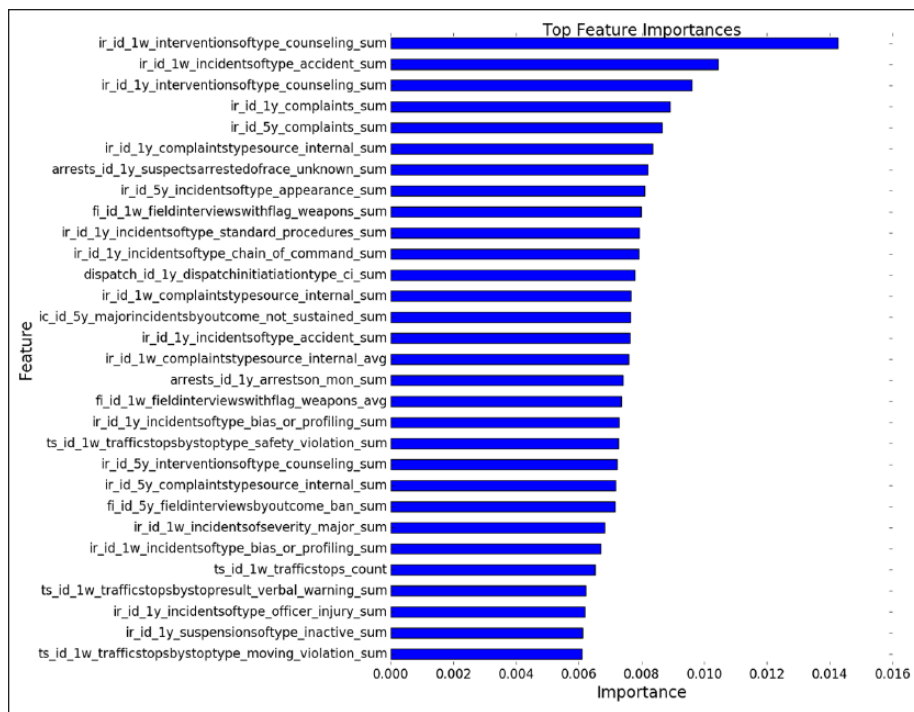
Figure 4. Feature importance for the best performing model.

non-officer-level factors may be able to find such officers so that interventions can be applied. Furthermore, these factors are overwhelmingly based on *behavioral characteristics of the officer, not demographic information*. While correlations are likely to exist between behavior and demographics, and causal factors may be extremely difficult to untangle, it is preferable to base policy decisions on attributes officers can remedy (behavior) as opposed to attributes they cannot easily change (weight).

Implementation and Next Steps

Given these results, CMPD has decided to replace its existing EIS with this data-driven system. CMPD has extensive experience building and deploying computer applications within the department, and our team has experience deploying machine learning systems in production environments. To ensure success, we have traveled to Charlotte, went on ride-alongs with officers, and met with numerous CMPD officers and staff, including the department leadership, IA investigators, and mid- and low-level supervisors. CMPD convened a supervisor focus group to evaluate model results, test the EIS interface, and discuss the department's EIS policy. Both sides participate in weekly or bi-weekly phone calls to ask questions, make suggestions, and monitor progress.

CMPD also convened an officer focus group, which included officers who were concerned about the project. In our experience, communication is key for the success of this work. There are two parts of the White House Police Data Initiative: opening some police data to the public and building data-driven early intervention systems. We are only involved in building EISs. We assured them that we would not make their private information (e.g., home addresses) available to the public and explained that the new EIS will help officers receive assistance while reducing false flags and the accompanying paperwork. They have been extraordinarily helpful, describing their work and suggesting variables to try. We expect the new EIS will be fully deployed by March, 2017.

The data-driven approach encourages departments to focus on the predictive accuracy of the EIS. We have talked with more than a dozen departments, and those with EISs rarely monitor their accuracy and adjust their models. We are providing training so the department can ensure that the model continues to perform as expected and, if it does not, try to fix it. In addition, CMPD's EIS interface gives supervisors the option to agree or disagree with each officer's EIS prediction and risk factors. The interface provides a way to capture some of information supervisors have that is not contained in the data, such as what is happening in an officer's personal life, and to obtain immediate feedback on the performance of the system. The data-driven EIS will include the supervisor information as variables, which should improve its accuracy over time.

CMPD plans to use the EIS in novel ways. First, the department leadership will look at the officers at the bottom of list. One possibility is to watch body-camera footage for unusually low-risk officers to see whether the department can learn what those officers are doing differently and whether it can be incorporated into CMPD's training. Second, CMPD will explore group-level predictions. Just as many departments provide specialized training and support for Special Weapons and Tactics (SWAT) teams, school resource officers, and other groups of officers, the department may provide group interventions. Group-level predictions would help them identify groups at risk. Finally, CMPD will begin to predict types of adverse events. The data-driven EIS predicts whether an officer will have any type of adverse event, but the machine learning approach can learn patterns for more specific outcomes, such as sustained complaints or officer injuries.

In terms of implementation, as always, the utility of the improved EIS will be mediated by social structures within the department. Perhaps, most importantly, supervisors using the EIS system should also be trained to treat the predictions by our model similarly. Instructing supervisors on how to understand the meaning of risk scores and how to interpret variables is important to our implementation approach.

Conclusion

The present work uses a machine learning approach to develop an EIS for flagging police officers who may be at high risk of involvement in an adverse interaction with the public. Our model significantly outperforms the existing system at the CMPD. Our model also provides risk scores to the department, allowing them to more accurately target training, counseling, and other interventions to officers who are at highest risk

of having an adverse incident. This allows the department to better allocate resources, reduce the burden on supervisors, and reduce unnecessary administrative work of officers who are not at risk.

Furthermore, our models provide insight into which factors are important in predicting whether an officer is likely to have an adverse event. We find that, largely, intuitive officer-level and neighborhood-level variables predict adverse events, but also that many variables the department had not yet considered also correlate with future adverse events. This information will, hopefully, allow this department, and potentially other police departments, to develop more effective early interventions for preventing future adverse events.

At a higher level, our goal is to take this system, developed for CMPD, and extend it to other departments across the United States. Work has begun with the Metropolitan Nashville Police Department and Knoxville Police Department, and the Los Angeles Sherriff's Department has committed to join. Other departments across the United States are considering joining. We have made our system open source for departments to build upon if they so choose.[5] A tool built across departments is especially important for small departments, which are unlikely to have enough adverse events to build reliable models.

Notes

1. Roughly, an indicator of the area the officer patrols and the time at which he or she patrols it.
2. Defined below.
3. http://mcmap.org/qol/
4. Randomization in the algorithm can lead to different lists of officers in the top 5%, if the algorithm is run multiple times. In our experiments, 10,000 trees increases the Jaccard similarity (percentage of overlap between lists) for two lists produced by the same algorithm to roughly 95%, a 15-percentage point increase over a 1,000-tree model.
5. https://github.com/dssg/police-eis

References

Arthur, R. (2015, December 15). How to predict bad cops in Chicago. *FiveThirtyEight, Politics*. Retrieved from https://fivethirtyeight.com/features/how-to-predict-which-chicago-cops-will-commit-misconduct/

Chapman, C. (2012). Use of force in minority communities is related to police education, age, experience, and ethnicity. *Police Practice & Research*, *13*, 421-436.

Commission on Accreditation for Law Enforcement Agencies. (2001). *Personnel early warning system*. Fairfax, VA: Author.

Davis, R. C., Henderson, N. J., Mandelstam, J., Ortiz, C. W., & Miller, J. (2005). *Federal intervention in local policing: Pittsburgh's experience with a consent decree*. U.S. Department of Justice, Officer of Community Oriented Policing Services. Washington D.C.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*, 3-42.

Goldstein, H. (1977). *Policing a free society*. Cambridge, MA: Ballinger.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Springer Series in Statistics). New York, NY: Springer.

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: Principles and practice*. OTexts: Melbourne, Australia. Available from http://otexts.org/fpp/. Accessed in April 2016

Jones, N. (2014). "The regular routine": Proactive policing and adolescent development among young, poor Black men. *New Directions for Child and Adolescent Development*, *143*, 33-54.

Manis, J., Archbold, C. A., & Hassell, K. D. (2008). Exploring the impact of police officer education level on allegations of police misconduct. *International Journal of Police Science & Management*, *10*, 509-523.

Shjarback, J. A. (2015). Emerging early intervention systems: An agency-specific pre-post comparison of formal citizen complaints of use of force. *Policing*, *9*, 314-325.

Shultz, A. (2015). *Early warning systems: What's new? What's working?* Center for Naval Analyses. Arlington, VA.

Sobol, J. J., Wu, Y., & Sun, I. Y. (2013). Neighborhood context and police vigor a multilevel analysis. *Crime & Delinquency*, *59*, 344-368.

Terrill, W., & Reisig, M. D. (2003). Neighborhood context and police use of force. *Journal of Research in Crime & Delinquency*, *40*, 291-321.

Timeline: Recent US Police Shootings of Black Suspects. (2015, April). *ABC News*. Retrieved from http://www.abc.net.au/news/2015-04-09/timeline-us-police-shootings-unarmed-black-suspects/6379472.

Topic:Police Brutality, Misconduct and Shootings. (2016, April). *The New York Times*. Retrieved from http://topics.nytimes.com/top/reference/timestopics/subjects/p/police_brutality_and_misconduct/index.html.

U.S. Commission on Civil Rights. (1981). *Who is guarding the guardians*. Washington, DC: Author.

U.S. Department of Justice. (2001). *Principles for promoting police integrity: Examples of promising police practices and policies*. Washington, DC: Author.

Walker, S. (2003). New paradigm of police accountability: The U.S. justice department pattern or practice suits in context. *Saint Louis University Public Law Review*, *22*, 3-52.

Walker, S., Alpert, G. P., & Kenney, D. J. (2001). *Early warning systems: Responding to the problem police officer*.U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Washington D.C.

White, M. D. (2002). Identifying situational predictors of police shootings using multivariate analysis. *Policing: An International Journal of Police Strategies & Management*, *25*, 726-751.

White, M. D., & Kane, R. J. (2013). Pathways to career-ending police misconduct an examination of patterns, timing, and organizational responses to officer malfeasance in the NYPD. *Criminal Justice and Behavior*, *40*, 1301-1325.

Worden, R. E., Kim, M., Harris, C. J., Pratte, M. A., Dorn, S. E., & Hyland, S. S. (2013). Intervention with problem officers an outcome evaluation of an EIS intervention. *Criminal Justice and Behavior*, *40*, 409-437.

Author Biographies

Jennifer Helsby is a postdoctoral researcher at the Center for Data Science and Public Policy at the University of Chicago where she works on the application of machine learning methods to problems in public policy. Previously, she was a Data Science for Social Good fellow and mentor.

Samuel Carton is a PhD candidate at the University of Michigan School of Information. Previously, he was a Data Science for Social Good fellow in 2015.

Kenneth Joseph is a postdoc at the Network Science Institute at Northeastern University and a fellow at Harvard's Institute for Quantitative Social Science. He completed his graduate work in the Societal Computing program in the School of Computer Science at Carnegie Mellon University. His research focuses on obtaining a better understanding of the dynamics and cognitive representations of stereotypes and prejudice, and their interrelationships with sociocultural structure.

Ayesha Mahmud is a PhD student in Demography at Princeton University, specializing in modeling the spread of childhood infectious diseases. In 2015, she was a Data Science for Social Good fellow.

Youngsoo Park is a postdoctoral researcher at the University of Arizona at Tucson. He was a Data Science for Social Good fellow in 2015 and received his Ph.D. in astrophysics from the University of Chicago in 2015.

Andrea Navarrete is a researcher at the University of Chicago's Centre for Data Science and Public Policy (DSaPP) where she works on using a data driven approach to public safety. Prior to DSaPP, she completed the Data Science for Social Good Fellowship (DSSG) where she worked with the Mexican government to improve response to citizen petitions. Previously, she has worked on analytics projects related to transportation and urban growth in Mexico.

Klaus Ackermann is a postdoctoral fellow at the University of Chicago's Center for Data Science and Public Policy. Previously he was a Data Science for Social Good fellow in 2015.

Joe Walsh is a research scientist at The University of Chicago's Center for Data Science and Public Policy. His work explores the application of machine learning and social science to applied problems in criminal justice, public safety, government transparency, public health, and transportation.

Lauren Haynes is currently Senior Project Manager at the Center for Data Science and Public Policy at the University of Chicago. Her prior experience includes 3.5 years as a consultant in Accenture's Technology Labs, IT Manager and Interim CIO at the Ounce of Prevention Fund,

and Product Manager at GiveForward. Lauren focuses on the use of Human Centered Design and usability for social good.

Crystal Cody is Computer Technology Solutions Manager at the Charlotte-Mecklenburg Police Department.

Estella Patterson is a police major at the Charlotte-Mecklenburg Police Department.

Rayid Ghani is the Director of the Center for Data Science and Public Policy at the University of Chicago. Previously he was the Chief Scientist at the Obama for America 2012 campaign.