Identifying Police Officers at Risk of Adverse Events*

Samuel Carton
University of Michigan
scarton@umich.edu

Jennifer Helsby
University of Chicago
jen@redshiftzero.com

Kenneth Joseph
Carnegie Mellon University
kjoseph@cs.cmu.edu

Ayesha Mahmud
Princeton University
mahmud@princeton.edu

Youngsoo Park
University of Arizona
youngsoo@email.arizona.edu

Joe Walsh
University of Chicago
jtwalsh@uchicago.edu

Crystal Cody
Charlotte-Mecklenburg Police Department
ccody@cmpd.org

CPT Estella Patterson
Charlotte-Mecklenburg Police Department
epatterson@cmpd.org

Lauren Haynes
University of Chicago
lnhaynes@uchicago.edu

Rayid Ghani
University of Chicago
rayid@uchicago.edu

ABSTRACT

Adverse events between police and the public, such as deadly shootings or instances of racial profiling, can cause serious or deadly harm, damage police legitimacy, and result in costly litigation. Evidence suggests these events can be prevented by targeting interventions based on an Early Intervention System (EIS) that flags police officers who are at a high risk for involvement in such adverse events. Today's EIS are not data-driven and typically rely on simple thresholds models based entirely on expert intuition. In this paper, we describe our work with the Charlotte-Mecklenburg Police Department (CMPD) to develop a machine learning model to predict which officers are at risk for an adverse event. Our approach significantly outperforms CMPD's existing EIS, increasing true positives by $\sim 12\%$ and decreasing false positives by $\sim 32\%$. Our work also sheds light on features related to officer characteristics, situational factors, and neighborhood factors that are predictive of adverse events. This work provides a starting point for police departments to take a comprehensive, data-driven approach to improve policing and reduce harm to both officers and members of the public.

1. INTRODUCTION

Recent high-profile cases of police officers using deadly force against members of the public have caused a political and public uproar [4, 5]. They have also highlighted and further encouraged tensions between the American 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 [13]. *Adverse events* between the police and the public thus come in many different 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 are predictive of adverse events [11, 6]. 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 not randomly distributed among officers or over time [11]. 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 responsively dealing with them after such an event occurs.

The work described in this paper was initiated as part of the White House's Police Data Initiative[1] launched based on President Obama's Task Force on 21st Century Policing. As part of this effort, we had discussions with several US police departments and it became clear that existing EISs were ineffective in their attempts to identify at risk officers. This paper describes our work with the Charlotte-Mecklenburg Police Department (CMPD) in North Carolina to use machine learning algorithms to improve their existing EIS.

CMPD's 1800 officers patrol more than 500 square miles encompassing more than 900,000 people. Over the last ten

*This work was started at the 2015 Eric & Wendy Schmidt Data Science for Social Good Summer Fellowship at the University of Chicago.

[1] https://www.whitehouse.gov/blog/2015/05/18/launching-police-data-initiative

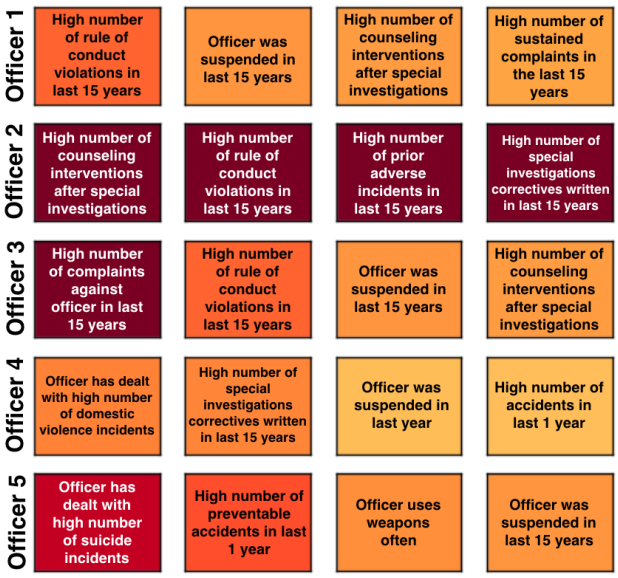

Figure 1: An illustration of five at-risk officers that will go on to have an adverse incident and their risk factors. The darker the red, the stronger the importance of that feature.

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. A supervisor then determines whether an intervention is appropriate. Several departments have adopted CMPD's system since it was built more than ten years ago [16]. To improve the current system, we focus on the following prediction task:

> Given the set of all active officers at time t and all data from time periods prior to t, predict which officers will have an adverse interaction in the **next year**.

We show that a random forest model with an extensive set of features significantly outperform the department's existing EIS. Specifically, our model shows a relative increase of $\sim$ 12% in true positive rate and a relative decrease of $\sim$ 32% in false negative rate over the existing EIS in a temporal cross-validation experiment. Unlike the existing system, our approach uses a data-driven approach and can thus be used to explore officer characteristics, neighborhood and environmental factors that are predictive of adverse events.

Figure 1 shows an illustrative chart that shows five officers and individual risk factors that them to have a high risk of an adverse event for five anonymous officers. Each officer in Figure 1 did indeed go 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.

In addition to factors we discover in our analyses of feature importances from this prediction task, we also provide an exploratory analysis of predictive features at the single event level to better understand situational factors that may play a significant role in scenarios leading to adverse events.

The system described here is the beginning of an effort that has the potential to allow police chiefs 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 them from happening in the first place.

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

- We apply, to our knowledge, the first use of machine learning toward prediction of adverse incidents from internal police department data.
- We show significant improvement over existing systems at flagging at-risk officers
- We take preliminary steps toward understanding the situational factors that lead to an adverse event.

2. EXISTING EARLY INTERVENTION SYSTEMS

A small minority of officers account for the majority of adverse events, such as citizen complaints or excessive uses of force [11, 6]. 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 US Commission on Civil Rights [1], the Commission on Accreditation of Law Enforcement Agencies [2], US Department of Justice [3], 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) to correct problematic policing require an EIS to be in place [19]. 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 [15].

Current EISs detect officers at risk of adverse events by observing a number of performance 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 as an officer working in the business district in the morning. More sophisticated systems flag outliers while conditioning on one or two variables, such as the officer's beat[2], but still fail to include many factors. For example, CMPD's indicators include complaints, uses of force, vehicle pursuits & 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

[2]Roughly, an indicator of the area the officer patrols and the time at which they patrol it

citizen complaints [20, 9], but it is unclear whether this decrease arises from a reduction in problematic behavior or from discouraging officers from proactive policing [23]. 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 [15].

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—which is good for the vendor but bad for the department. 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 are limited to binary flags instead of risk scores. Risk scores enable the agency to rank people/facilities/etc. by risk, to explicitly choose tradeoffs (e.g. precision vs. recall), and to allocate resources in a prioritized manner.

A machine learning system would be able to alleviate many of these issues. With respect to customization, machine learning models can be easily retrained on new data and with new features. Furthermore, given the volume of features and feature interactions that can be used within a machine learning model, parameters are sufficiently complex that the system cannot be easily gamed. Importantly, such models return control to the department, allowing its leaders to choose the right mix of accuracy and interpretability. 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 American 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.

3. POLICE MISCONDUCT

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

More educated police officers, particularly those with four-year college degrees, tend to have fewer complaints and allegations of misconduct compared to officers with less education [14, 22, 8]. In a study of misconduct in the New York Police Department, White and Kane [22] 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 [21] 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 [17]. Research on citizen characteristics has, moreover, been limited due to lack of publicly available data.

Finally, neighborhood features have also been studied as a potential predictor of police misconduct. Sobol [17] found that incidents in high-crime neighborhoods have a greater likelihood of ending in interrogation, search and/or arrest. Similarly, Terrill and Reisig [18] found that police officers were more likely to use higher levels of force in disadvantaged and high-crime neighborhoods.

Our models incorporate features at each of these levels of analysis, finding that predictors 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 features correlated with officers at risk of adverse events.

Database	Num. Records	Time Window
Internal Affairs	20K	2002-Now
Dispatch Events	14M	2003-Now
Criminal Complaints	959K	2005-Now
Citations	946K	2006-Now
Traffic Stops	1.6M	2002-Now
Arrests	350K	2005-Now
Field Interviews	180K	2003-2009
Employee Records	20K	2002-Now
Secondary Employment	651K	2009-Now
Training	1.4M	2001-Now
Existing EIS	14K	2005-Now

Table 1: Description of the types of data used, as well as the number of records and the time window over which we have data of that type

4. DATA DESCRIPTION

The data for this work consists of almost all employee information and event records collected by CMPD to manage its day-to-day operations. Certain information, such as employee names, 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 dataset, summarized in Table 2, 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.

4.1 Internal Affairs Data

Internal Affairs (IA) records contain the information about adverse events that we use as our outcome variable. Every IA record pertains to a single officer. When a department employee or member of the public files a complaint or when

Event	IA Ruling
Vehicle Accidents	Preventable
Use of Force	Unjustified
Raid and Search	Unjustified
Citizen Complaint*	Sustained
Officer Complaint*	Sustained
Pursuit	Unjustified
Injuries	Unjustified
Discharge of Firearm	Unjustified
Tire Deflation Device	Unjustified

*Minor violations excluded

Table 2: The types of events within the IA database that we define as representative of an *adverse event*

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*[3] 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.

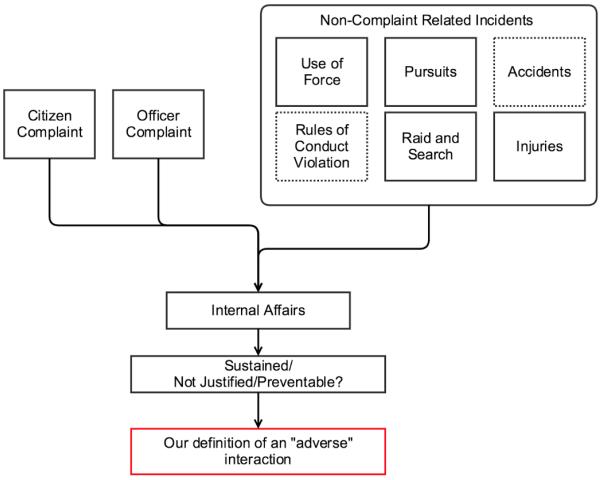

Figure 2: The Internal Affairs process and our definition of an adverse incident.

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 define records with *not justified*, *preventable*, and *sustained* dispositions to define the class of adverse events, with exceptions for a number of internal complaints that we consider less egregious, such as misuse of sick leave.** These data serve as the positive class for our dependent variable. 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.

[3]defined below

Notably, we proceed with the assumption that the Internal Affairs (IA) data reasonably represent 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 complete 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.

4.2 Other Data

4.2.1 *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.

4.2.2 *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.

4.2.3 *Citations*

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

4.2.4 *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 socio-demographic profile.

4.2.5 *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.

4.2.6 Field Interviews

A "field interview" is the broad name given by CMPD for any event in which a pedestrian is stopped and/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 result of a traffic stop. Records contain temporal and spatial information as well as information about the demographics about the interviewed person.

4.2.7 Employee Records

The department's employee information includes demographic information on every individual employed by the department, including those that have retired or been fired. The data includes officer education levels, years of service, race, height, weight, and other persistent qualities of officers.

4.2.8 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.

4.2.9 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.

4.2.10 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 an 90 day period) and the selected intervention for each flag, which can include training and counseling.

4.2.11 Neighborhood

In addition to the data provided by CMPD, we also use publicly available data from 2010 and 2012 neighborhood quality-of-life studies[4] to understand the geospatial context of CMPD events. These studies collect data on many neighborhood features including Census/ACS data on neighborhood demographics and data on physical characteristics, crime, and economic vitality.

4.3 Data Limitations

In addition to the potential bias discussed above, the dataset 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 feature generation.

[4] http://mcmap.org/qol/

5. METHODS

The goal of the EIS is to predict which officers are likely to have an adverse event in the near future. We formulate it as a binary classification problem where the class of interest is whether a given officer will have an adverse event in a given period of time into the future. In discussions with CMPD and in consideration of the rareness of adverse events, we decided that one year was an appropriate prediction window. Efforts were chiefly geared towards the extraction of these features - in total 432 features were used. For modeling, we tried a variety of model types, including AdaBoost, Random Forests, Logistic Regression, and Support Vector Machines. Random searches over a standard hyperparameter space using 3-fold cross-validation were used to tune each model. Below, we discuss our feature extraction process and how models were evaluated.

5.1 Feature Generation

We generated features based on our expertise as well as on discussions with experts at the Charlotte-Mecklenburg Police Department. Patrol officers, Internal Affairs investigators, members of our officer focus group, and department leadership suggested features that varied across the officer- and neighborhood- levels of analysis. We explore situational factors in Section 7.

At the officer level, we generate behavioral features by aggregating the record of incidents by each officer, establishing a behavioral history. The simplest features are frequencies 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). Broad incident classes we track include arrests, traffic stops, field interviews, IA cases, and external employment.

Notably among incident sub-types, 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 incorporated features describing the number of credit hours of trainings officers had in topic areas of relevance: less-than-lethal weapons training, bias training, and physical fitness training.

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

We include the existing EIS thresholds as features in our model. These EIS flags will occur if a threshold number of adverse events occur within a specific timeframe, e.g. 3 uses of force within 90 days, and similarly for other potential warning signs such as complaints and sick leave use.

Finally, we include neighborhood features to capture specific information about the areas where officers patrol. For

Metric	Existing EIS	Improved EIS	Percent Change
True Positives	43	48	+12%
False Positives	624	427	−32%
True Negatives	802	999	+25%
False Negatives	40	35	−8%

Table 3: Comparison of model performance between the existing threshold-based EIS and the improved predictive EIS developed in this work.

example, we included the 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.

5.2 Model Evaluation

We validate our models using temporal cross validation [12], meaning that if, for example, predictions were being made for adverse events in the years 2010-2012, we train our models on data from periods before 2010. With our data ranging from 2009 to 2015, we perform multiple evaluations over the data and aggregate them to come up with the final statistics. For each evaluation, we use precision (percent of officers flagged who actually have an adverse event) and recall (percent of officers with adverse events who are flagged) at various probability (or risk score) thresholds as outcome metrics. We compare various versions of our models and feature sets to each other as well as to a random baseline, to a classifier that exactly replicates the current EIS, and to a logistic regression baseline model using only the officer age, sex, race, years of experience and days since last adverse event as features.

6. RESULTS

In this section we discuss results in terms of performance on the officer-level prediction problem as well as an analysis of highly predictive features.

6.1 Predictive Performance

At the officer-level, about $\sim 8-9\%$ of officers will end up having an adverse event of some type in a one year period. The best binary classification model to predict these events was a Random Forest with 50 estimators. Table 3 shows how our model compares with the EIS baseline in terms of false positives, false negatives, true positives, and true negatives. Our results show that moving beyond the current threshold system and using a broader set of data with more complex models improves the existing system. Our best performing model is able to flag 12% more high-risk officers (true positives), while flagging 32% fewer low-risk officers (false positives) compared to the current system. We show the precision-recall curves for the officer-level prediction problem in Figure 3.

6.2 Feature Analysis

Figure 4 shows the features with the largest feature importances in our best performing random forest model. The most predictive features of the model 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

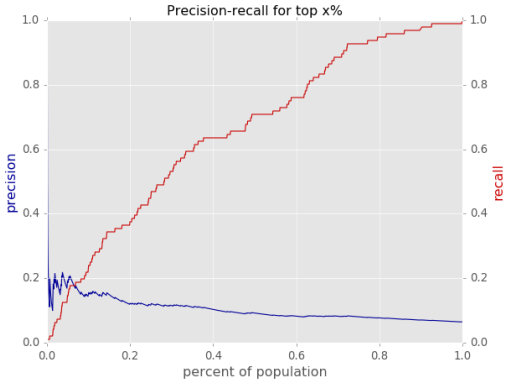

Figure 3: Precision-recall curves for the Random Forest model.

to engage in another such event in the future. This is fairly typical in behavioral prediction tasks.

Such indicators are complex and overlie a variety of causal factors - for example, officers who are in areas of higher rates of violent crime are more likely to use force because of the area they patrol and perhaps not because of any inherent tendencies. However, two caveats to this notion are in order. First, significant controls at the neighborhood level exist within the model. Such controls have an impact on prediction - for example, vacant land area rates are a significant predictor of officer risk. Second, indicators such as the rates of prior adverse incidents and sustained complaints indicate cases where IA officials previously found officers to be at fault over and above these increased risk rates.

Combined, these observations provide support for the idea that a subset of officers are at particular risk for adverse events, and that an EIS which controls for non-officer level factors may be able to find such officers so that interventions can be applied. Further, these factors are 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 things officers can remedy (behavior) as opposed to things they cannot easily change.

To maximize the insight gained from the most prominent features, it is ideal to have information on the *directionality* of these features, i.e. whether the changes in one of the features correlate positively or negatively with the corresponding change in the predicted risk score. Such information would clarify how a feature moves the trained model, thereby allowing for a deeper understanding of the underlying phenomenon. Traditional approaches to feature importance in random forest models do not, however, allow us to infer the direction of a feature's effect. In order to address this issue, we perform a Monte Carlo sampling of our model's risk score surface to estimate the conditional distribution of risk scores, or the "risk score curve," on each feature.

We begin with 100,000 Monte Carlo samples generated by drawing from a uniform distribution in the feature space spanning the entire range of feature values in the training data. To analyze the risk score curve of a feature, we divide the generated samples into 50 bins ranging from the minimum to the maximum value of the feature of interest

Figure 4: Feature directionalities from the officer-level random forest model

found in the dataset. Finally, within each bin, the mean and the standard deviation of the risk score distribution are calculated. In Figure 5, we present the risk score curve for the number of uses of force in traffic stops over the last 15 years. It is notable that a sharp transition, or shift, in the risk score distribution occurs around seven traffic stops. This sharp transition aligns with the behavior one would expect from a random forest model, where risk scores are determined via binary selection criteria that act as sharp distributional "switches".

To construct a metric for directionality from the estimated risk score curve, we use the mean risk score difference between the first and last bin. While this metric is arguably less robust for features that undergo multiple transitions of conflicting directions, it is a good first-order proxy for the directionality of the most prominent features. Further, in assessing the resulting data, such multiple transitions were rare in the feature set used for the present work.

After estimating the risk score curve for every feature used, we present in Figure 4 the most prominent directionalities we found in our model. It is worth noting that while there is a strong correspondence between features with high importances and high directionalities, it is not an exact match. This is because the definition of feature importance in random forests depends not only on the strength of the directionality of a feature, but also on the exact configuration of the trees within the forest. We are actively looking at other ways of determining feature importance, such as using additive models [7].

Figure 5: Risk score curve for the number of uses of force in traffic stops over the last 15 years.

7. DISPATCH-LEVEL PREDICTION

As an exploratory exercise to better understand situational, near-term factors that may have an impact on officer risk of adverse events, we attempt to predict which dispatch calls are likely to result in an adverse event. Each dispatch call record in the data contains data that includes time, location, type of call, officers dispatched, and priority (or urgency) of the situation. These environmental factors of a given event could play a significant role in determining whether an event "turns adverse", in addition to the characteristics of officers involved. Furthermore, a history of dispatch calls can be constructed for each officer, from which a general pattern of dispatches leading to an eventual adverse event can be found. For example, overworked officers at the end of a long shift may be more likely to be involved in an adverse event, and this analysis allows us to discern whether

such patterns exist.

To make predictions at the dispatch-level, we use most of the features generated for the officer-level experiment. To these we add features of the dispatch event itself, such as its priority level, features of medium and short-term officer stress, such as how many consecutive days the officer had been on duty at the time of the dispatch, and features of the location in which the dispatch takes place. In total, we examine 359 features at the dispatch level.

For this task, there is no existing baseline method analogous to the EIS. Therefore, all comparisons are against a random baseline. Further, and most importantly, adverse events are extremely rare: 1 in 10,000 dispatches end in any type of adverse incident in our dataset. As an exploratory analysis, we subset the data to a ratio where feature analysis can be performed. This means that model performance *should not be expected to hold in realistic settings.*

The positive examples for this prediction task consist of every dispatch from CMPD's database that can be linked to an adverse event. These 929 positive examples are contrasted against 8,361 negative examples (i.e. "non-adverse" events) drawn randomly from the database, for a 10%-90% balanced training set of 9,290 examples. We then split the data temporally, training on all adverse events prior to 2013, and testing on those following 2013.

To understand what types of feature lend utility to this prediction model, we compare performance of different feature subsets. Notable subsets we examine include dispatch features, such as the priority level of the dispatch (1 to 9) or the typecode assigned to it by the dispatcher (e.g.SUSP-SCN for suspect-on-scene), and medium-to-short-term officer stress features, such as how many hours the officer has been on duty at the time of the dispatch.

Figure 6 shows the performance of a tuned random forest on predicting whether dispatch events will result in adverse interactions between the involved officer and a citizen. The full model achieves an F1 of 0.478 with respect to the positive class, significantly better than the 0.18 that would result from random guessing.

Features of the dispatch event itself dominate the model. Used alone, they achieve comparable performance to the full model. Removed from the dataset, they reduce model performance to indistinguishable from random. This suggests that immediate situational factors outweigh longer-term officer- or location-level factors in determining when a dispatch is likely to result in an adverse event.

Figure 7 examines which features are used to the greatest cumulative effect in reducing sample impurity in the random forest model. The clear outlier is travel time, which appears to have a major impact in predicting adverse outcomes. Other significant features include the JST-OCC dispatch typecode, indicating an event that has just occurred, and the career arrest rates, both discretionary and overall, of the officer involved in the dispatch.

Figure 8 further examines feature importance by using the same Monte Carlo sampling method employed in figure 4. Travel time is found to have a positive sign, meaning that longer travel times are associated with a higher risk of adverse outcomes in the model. The dominant positive feature by this measure of importance is the REPT-OFC dispatch typecode, indicating that the situation being addressed by the dispatch was reported by the responding police officer. Similar, and also positively contributing, are

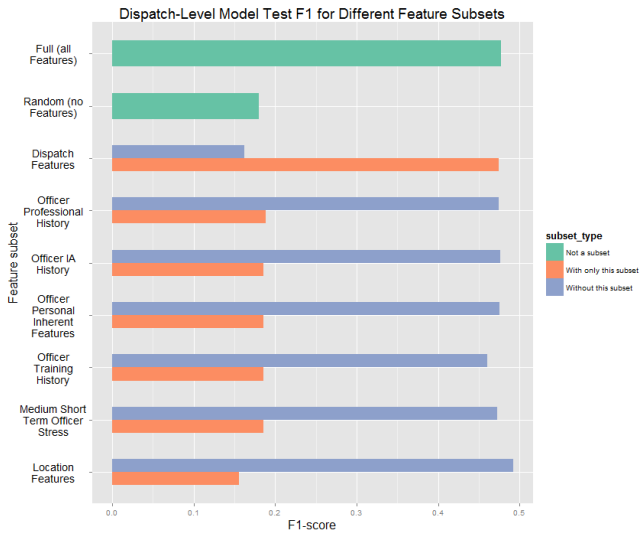

Figure 6: Comparison of model performance (f1-score w/respect to positive class) of feature subsets on dispatch-level adverse-outcome prediction

the OFC-INVL (officer involved) typecode and OI (officer-initiated) dispatch type. Other contributing features include the suspect-on-scene typecode, the number of hours the officer has been on duty, and two features associated with how frequently that officer makes discretionary arrests

Features that contribute negatively to the risk of adverse outcomes include the height and weight of the officer, and the number of days since their last discretionary arrest. Wealthier neighborhoods with a higher age of death are associated with fewer negative outcomes, though interestingly, so are those with a greater number of minor nuisance violations.

Taken together, these results seem to reinforce the conclusion that situational factors are largely, though not exclusively, predictive of adverse outcomes at the individual dispatch level. "Hot" dispatches initiated by officers themselves (as opposed to citizens by way of 911 calls), seem more likely to end in adverse outcomes. Indicators of heightened officer stress (hours on duty) and aggressive policing style (discretionary arrest rate), seem to also have a positive impact on the risk of adverse outcomes.

8. IMPLEMENTATION AND NEXT STEPS

Next steps include implementation and effect of interventions. There are several ways the risk scores could be used by a police department. However, our primary goal is to develop individually tailored interventions to ensure that each officer receives appropriate training and support. In addition, the risk scores enable the prioritization of resource allocation to the officers that are considered most at-risk.

We are exploring using our model to develop interventions for groups of officers. When risk scores are aggregated over groups defined by unit or division, we find that some divisions and units have a significantly higher risk than others. These divisions and units may benefit from additional group interventions such as group trainings to lower their risk.

In terms of implementation, as always, the utility of the improved EIS will be mediated by social structures within

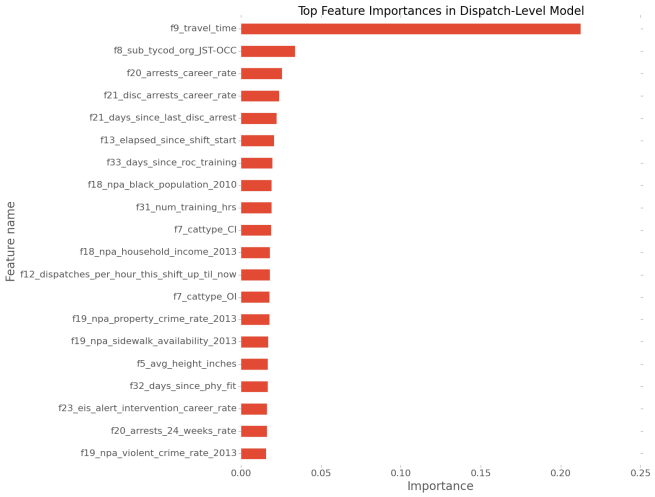

Figure 7: Conventional unsigned feature importances in dispatch-level random forest model

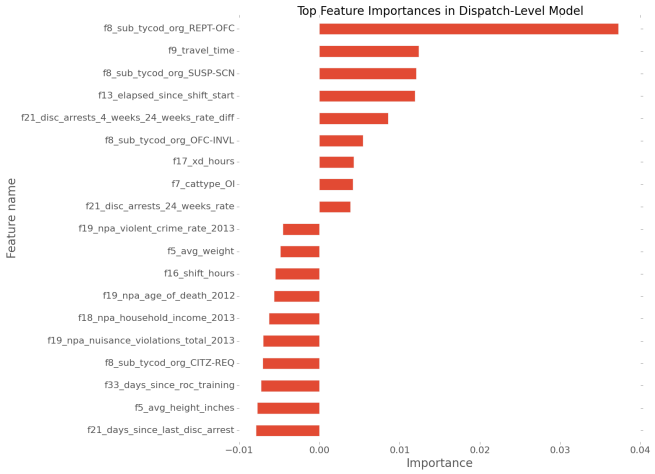

Figure 8: Inferred signed feature importances in dispatch-level random forest model

the department. Perhaps most importantly supervisors using the EIS system should also be trained to treat model results similarly. Instructing supervisors on how to understand the meaning of risk scores and how to interpret features will be an important avenue of our implementation approach.

Our dispatch-level models take the first steps toward predictive risk-based dispatch decisions, where an officer who is at higher risk of an adverse incident for that dispatch can potentially be held back and a different officer, at a lower risk score, can be dispatched. For example, in June 2015 a police officer in Texas, Cpl. Eric Casebolt, pulled his weapon on children at a pool party after responding to two suicide calls earlier that shift [10]. Most police departments would like to avoid these situations by dispatching low-risk officers to calls. Risk-based dispatching could enable improved dispatching to match officers and dispatches while minimizing risk of harm to the public and the officer.

Finally, future work will focus on finding the appropriate balance between actionability, transparency, interpretability, and resistance to gaming by officers. A dashboard can help strike a balance between these concerns and communicate model results in easy-to-use and actionable formats, which we are currently developing for use by the CMPD. The proposed system will provide the top feature importances, which will enable officers in the department to understand what factors are typically correlated with adverse incidents without providing a recipe for those that wish to game the system.

9. CONCLUSION

The present work uses a machine learning approach to develop an Early Intervention System for flagging police officers who may be at high risk of involvement in an adverse interaction with a member of the public. Our model significantly outperforms the existing system at the Charlotte-Mecklenburg Police Department (CMPD). Our model also provides risk scores to the department, allowing them to more accurately target training, counseling, and other interventions toward officers who are at highest risk of having an adverse incident. This will allow the department to better allocate resources, reduce the burden on supervisors, and reduce unnecessary administrative work of officers who were not at risk.

Further, 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 features are predictive of adverse events, but also that many features the department had not yet considered are also correlated 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.

To explore the immediate situational factors associated with adverse events, we also engaged in an exploratory analysis at the individual dispatch event level. Results suggest that features of a particular dispatch may be highly predictive of whether or not a dispatch will result in an adverse outcome relative to officer-specific features. Future work will be focused on addressing how to utilize these features more effectively.

At a higher level, our goal is to take this system, developed for CMPD, and extend it to other departments across the US. We already have commitments from Los Angeles Sherrif's Department and Knoxville Police Department to work with us to extend this system. Several other departments across the US are also in discussions with us. 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. We are also implementing the system on CMPD It systems and monitoring the model's performance one year in the future, from July 1, 2015, which is the last day of data we received, to June 30, 2016.

Finally, we are discussing an intervention pilot in partnership with CMPD. Predicting which officers will have adverse events will only be impactful if it is possible to design interventions to prevent those events. Similarly, we realize that while intervention may reduce adverse events between the police and the public, such interventions are only a part of

[5] `https://github.com/dssg/police-eis`

a larger approach to dealing with the complex web of cognitive, interactional, social, and institutional factors affecting the relationship between the police and the public. We are hopeful that work at the intersection of data science, social science and the practice of policing can someday help to advance the work being done in these contexts as well.

10. ACKNOWLEDGMENTS

We thank the Eric & Wendy Schmidt Data Science for Social Good Fellowship for generously supporting this work. We also thank the leadership and officers of the Charlotte-Mecklenburg Police Department for sharing data, expertise, and feedback for this project as well as the White House Office of Science and Technology Policy for their help and support.

References

[1] *Who is Guarding the Guardians?* United States Commission on Civil Rights, Washington, DC, 1981.

[2] *Personnel Early Warning System.* Commission on Accreditation of Law Enforcement Agencies, Fairfax, VA, 2001.

[3] *Principles for Promoting Police Integrity.* United States Department of Justice, Washington, DC, 2001.

[4] Timeline: Recent us police shootings of black suspects. `http://www.abc.net.au/news/2015-04-09/timeline-us-police-shootings-unarmed-black-suspects/6379472`, 2015.

[5] Topic: Police brutality, misconduct and shootings. `http://topics.nytimes.com/top/reference/timestopics/subjects/p/police_brutality_and_misconduct/index.html`, 2016.

[6] R. Arthur. How to predict bad cops in chicago. `http://fivethirtyeight.com/features/how-to-predict-which-chicago-cops-will-commit-misconduct/`, 2015.

[7] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1721–1730, New York, NY, USA, 2015. ACM.

[8] C. Chapman. Use of force in minority communities is related to police education, age, experience, and ethnicity. *Police Practice and Research*, 13(5):421–436, 2012.

[9] R. C. Davis, N. J. Henderson, J. Mandelstram, C. W. Ortiz, and J. Miller. *Federal Intervention in Local Policing : Pittsburgh's Experience with a Consent Decree.* Office of Community Oriented Policing Services, Washington, D.C, 2003.

[10] M. Fernandez. Texas officer was under stress when he arrived at pool party, lawyer says. *New York Times*, June 10 2015.

[11] H. Goldstein. *Policing a Free Society.* Ballinger Pub. Co, Cambridge, Mass, 1977.

[12] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice.* OTexts, 2014.

[13] N. Jones. âĂIJthe regular routineâĂİ: Proactive policing and adolescent development among young, poor black men. *New directions for child and adolescent development*, 2014(143):33–54, 2014.

[14] J. Manis, C. A. Archbold, and K. D. Hassell. Exploring the impact of police officer education level on allegations of police misconduct. *International Journal of Police Science and Management*, 10(4):509–523, 2008.

[15] J. A. Shjarback. Emerging early intervention systems: An agency-specific pre-post comparison of formal citizen complaints of use of force. *Policing*, 9(4):314–325, mar 2015.

[16] A. Shultz. Early warning systems: What's new? what's working?, 2015.

[17] J. J. Sobol, Y. Wu, and I. Y. Sun. Neighborhood context and police vigor a multilevel analysis. *Crime & Delinquency*, 59(3):344–368, 2013.

[18] W. Terrill and M. D. Reisig. Neighborhood context and police use of force. *Journal of Research in Crime and Delinquency*, 40(3):291–321, 2003.

[19] S. Walker. The new paradigm of police accountability: The u.s. justice department "pattern or practice" suits in context. *Saint Louis University Public Law Review*, 22(3):3–52, 2003.

[20] S. Walker, G. P. Alpert, and D. J. Kenney. *Early warning systems: Responding to the problem police officer.* US Department of Justice, Office of Justice Programs, National Institute of Justice, 2001.

[21] M. D. White. Identifying situational predictors of police shootings using multivariate analysis. *Policing: an international journal of police strategies & management*, 25(4):726–751, 2002.

[22] M. D. White and R. J. Kane. 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(11):1301–1325, 2013.

[23] R. E. Worden, M. Kim, C. J. Harris, M. A. Pratte, S. E. Dorn, and S. S. Hyland. Intervention with problem officers: An outcome evaluation of an EIS intervention. *Criminal Justice and Behavior*, 40(4):409–437, oct 2012.

.