Nayar Prize II, Phase I Quarterly Progress Report (*Q4*)
October 2017

Project: A Data-Driven Crime Prevention Program
Team: Miles N. Wernick, Lori Andrews, Yongyi Yang
Students: David Haro Alonso, Pauline Panayi, Harshini Tippareddy, Raven Zeer
Elgin Police Department: Bill Wolf, JoAnn Stingley, Pastor Robert Whitt

Progress Summary of Nayar Prize II, Phase I

Project overview:

The objective of our project is to design, implement, and deploy a flexible, new model for crime prevention that can be translated to a wide array of communities in the United States and beyond, thereby achieving far-reaching societal impact. The project is producing new developments in predictive modeling technology based on machine learning, along with a legal-ethical framework for appropriately employing this technology to crime prevention in a way that respects privacy rights and achieves acceptance by the community. The project is a partnership with the Elgin (Illinois) Police Department, which we selected because of its exceptionally progressive culture, positive relationship with its community, and willingness to devote substantial effort to work closely with us in a genuine partnership.

Our crime-prevention program parallels current directions in the field of preventive medicine, where it has been recognized that intervention strategies should shift from a reactive stance to one that is proactive. Intervention strategies in preventive medicine often begin with statistical models that estimate an individual's risk for developing disease based on various risk factors (such as smoking), and then deploy treatment (such as smoking cessation) to reduce future risk.

In an analogous approach, we are developing risk models, founded on machine learning technology, that assess an individual's future risk of becoming a victim or perpetrator of crime. In our intervention strategy, individuals identified by the model to be at high risk will be invited to work with social workers who will help the individual to identify which social services programs may be most helpful to them (these may include job training, drug rehabilitation, and educational opportunities), and then follow up with the individual consistently to ensure success. As in preventive medicine, the aim is to reduce the individual's future risk by modifying the circumstances that gave rise to that risk.

The overall goal of our project is to help people to better their lives, thereby benefiting them and their families and ultimately leading to a safer community. Our objective is to demonstrate an approach to crime prevention that will serve as a model to other communities nationwide and throughout the world. We will disseminate knowledge gained in the project through scientific publications; law review articles; training for police officers, community leaders, and judges; media interviews and/or op-ed pieces; and interactions with others interested in using risk modeling in their social services activities.

Executive summary of progress report:

In Year 1, the project proceeded according to the steps we outlined in our original proposal. We began by assessing the needs of the Elgin community, working with a team of representatives of the Elgin Police Department (EPD), including the deputy chief of police and other subject-matter experts, the department's head social worker, and a local pastor, who acts as a liaison to the community and local churches. We discussed and selected strategies in a collaborative and iterative process involving our legal and technical experts, our students, and the EPD. We conducted a detailed legal and ethical analysis of relevant case law to guide the development of the risk model and the intervention.

We undertook technical development of several preliminary versions of the machine learning risk model and, after extensive consultation with the EPD, we selected one of the versions to become the basis for the first pilot phase of interventions. We performed a thorough statistical analysis of the risk model, not only to test its accuracy, but also to ensure that no racial or ethnic bias was present in the results. We also evaluated the risk model subjectively by computing risk assessments for individuals in the past to see whether experts from within the EPD would agree that the assessments are reasonable and insightful. This step was a resounding success, with experts reporting that the model produced recommendations that pointed to individuals who later were involved in very serious crimes, or who had been successfully assisted by previous social service outreach programs. In a fully collaborative effort, the EPD and our team developed a Standard Operating Procedure (SOP) document to define the goals and procedures for implementation of the intervention program. We have thoroughly reviewed the intervention concept and SOP with the head social worker and pastor, and the basic logistics have been decided. The pastor has shared our program's plans at a meeting of local church leaders to obtain feedback and community buy-in. His deep familiarity with the local community and his ability to talk to at-risk individuals in a way that resonates with them will make him an invaluable resource to engage them in seeking the help that they need. At this writing, the SOP document is undergoing a final approval process by the City of Elgin, and the pilot interventions will begin as soon as that process is complete. In addition to our research work in Year 1, we began disseminating knowledge gained and joined the national dialog on crime. We participated in media interviews relating to crime prevention (e.g., *60 Minutes* and *The New York Times*); we gave invited talks to classes at DePaul and Illinois Tech; we initiated discussion with other interested institutions and programs, and we began preparation of academic publications based on our work.

The following sections provide details of our progress in Year 1.

Collaboration with the Elgin Police Department (EPD):

Background
Miles Wernick, P.I. of this project, has been actively researching the concept of using data analysis as a tool for crime prevention for several years. Based on this work, he has been approached by several police departments asking him to collaborate on risk-assessment models for their communities. We chose to work with the EPD because of its history of innovative, culturally sensitive, community-based programs to prevent crime and create a safer, more productive community. Our selection proved to be even more fortuitous than expected, with the EPD fully incorporating our recommendations at every step of the way and showing great sensitivity and progressive thinking.

In the late 1980s, Elgin's crime problem was so severe that two documentaries were made about it. In response, the EPD adopted several strategies to prevent crimes. For example, EPD began an officer-in-residence program in which police officers were given housing in the most dangerous parts of the city. Initially, the program posed great risk to the officers involved (their houses were set afire and their cars blown up), but ultimately, the individual police officers became integrated into their local communities. Seeing first-hand what was happening on the streets and in the schools, they were able to help develop programs to serve the needs of community members.
The previously dangerous neighborhoods that officers volunteered to live in have now become much safer. But Elgin faces other problems—crimes by the homeless and those who are mentally ill, gang problems (with all the gangs in Chicago having members in Elgin as well), and other serious crimes.

For the past 26 years, the EPD has maintained a social work division that runs programs in the high schools, matches parolees with social services programs that can benefit them (such as drug rehabilitation and GED programs), and provides extensive aid in domestic violence cases which are attuned to the specific needs of the individuals, such as when the mortgage crisis triggered increased domestic violence.

In working with the EPD during this past year, we have been impressed by their commitment to provide relevant social services—*with consistent follow- through*—to the people that our assessment will identify as being at risk. In addition, the EPD worked with us to develop a standard operating procedure (SOP) entitled "Social Services Crime Prevention Program Procedures." This SOP meets the legal and ethical guidelines that we developed through our legal-ethical analyses, and it provides

CODE	DESCRIPTION
Homicide	
0110	1st-degree murder
0115	Homicide, unborn child
0120	Voluntary manslaughter, unborn child
0130	2nd-degree murder
0141	Involuntary manslaughter
0142	Reckless homicide
0150	Justifiable homicide
0160	Concealment of a homicide death
0165	Involuntary manslaughter & reckless homicide of an unborn child
0170	Drug induced homicide
0190	Solicitation for murder or murder for hire
Criminal sexual assault	
0260	Criminal sexual assault
0261	Aggravated criminal sexual assault
0262	Forcible sodomy
0280	Predatory criminal sexual assault of a child
0281	Criminal sexual assault with an object
Robbery	
0310	Armed robbery
0320	Robbery
0325	Vehicular hijacking
0326	Aggravated vehicular hijacking
0330	Aggravated robbery
Battery	
0410	Aggravated battery
0460	Battery
0470	Reckless conduct
0475	Battery of an unborn child
0480	Heinous battery
0485	Aggravated battery of a child
0486	Domestic battery
0487	Aggravated battery of an unborn child
0488	Aggravated domestic battery
0491	Aggravated stalking
0495	Aggravated battery of a senior citizen
Assault	
0510	Aggravated assault
0560	Assault

Table 1.

clear rationales and processes for the administration of the program, its interaction with participants, its   transparency, and its obligations to the community (See Appendix A). We are confident that, going forward, the EPD will provide the access we need to assess the social services interventions and refine the approach as the process proceeds.

Understanding Elgin's needs and resources: The EPD's motivation to initiate the project arises from the recognition that a relatively small fraction of Elgin's population accounts for most of the serious crime activity; thus, by helping even a few people to reduce their involvement in crime, the EPD feels that the program can have a significant impact on the community.

In contrast to Chicago, which is suffering from a unique epidemic of gun violence, the EPD explained to us that Elgin's crime problems are more typical of small- to medium-sized U.S. cities, such as those found in parts of the country where the local economy is not as vibrant as it once was. Thus, whereas our team focused its prior work with the Chicago Police Department on gun violence and homicides, we determined early on that Elgin would be better served by an approach that considers a broader array of serious crimes. The EPD's deputy chief and its subject-matter experts discussed the project goals and proposed a list of crimes that they consider to fall into this category (See Table 1).

The EPD feels that our approach will help them to identify individuals in a thorough, systematic, and data-driven way, whereas currently only a subset of at-risk individuals are referred for outreach (such as when police or social workers happen to come across these individuals through their daily activities). While their current outreach efforts are successful, the EPD feels that it is important to reach out to individuals who are at extreme risk, but whose situations may be overlooked because they are simply unknown to the social workers.

An individual may have a spectrum of needs based on his or her circumstances, with assistance depending in part whether the individual is a victim or arrestee. However, we decided that the judgment as to what assistance would be most appropriate should be decided on a case-by-case basis through a careful discussion between the individual and a social worker. Therefore, we have decided that the risk model would not seek to differentiate victims from arrestees. Such an approach also makes sense in view of the fact that those at risk for involvement in serious crimes often become both victims *and* perpetrators within a given time period.

The EPD has offered substantial personnel resources to the project. Deputy Chief Bill Wolf, EPD's principal point of contact for the project, will be responsible for top-level policy-making and discussions about the project direction. The EPD will also involve several high-ranking police officers, who have first-hand interactions with at-risk individuals, and who have expert knowledge of the neighborhoods and challenges. The EPD has also assigned a very capable member of their information technology team to assist with access to relevant crime records.

On a day-to-day basis, the most important EPD personnel in the program will be a team of social workers, led by JoAnn Stingley, social services supervisor, and Pastor Robert Whitt, community liaison. Stingley has been applauded in the media as an "unsung hero"[1] for her selfless community activities, leading fundraisers that support many of the social programs in Elgin, in addition to being the head social worker of the EPD. Whitt is a member of Elgin's Human Rights commission and is a well-known community leader who engages in efforts to promote healthy communications between police and the public and to foster racial harmony,

such as through the community meeting he organized to help people of various backgrounds to understand the Black Lives Matter Movement.[2]

We have found the EPD team to be very impressive, showing a remarkable degree of sensitivity that is in stark contrast to the negative public image of some police departments. The EPD places a great deal of emphasis on community-oriented policing, social work, and addressing the underlying needs of people involved in crime, rather than treating policing as a process of simply arresting people. Our conversations with various Elgin police officers have demonstrated that this progressive culture extends down to the rank-and-file officers.

Legal and ethical research:

To provide a sound basis for development of our risk model and the associated social services intervention program, we conducted a thorough legal and ethical analysis relating to the uses of algorithms in relation to crime. A disturbing trend is emerging in which algorithms are being used, or at least considered, in ways that we consider inappropriate, including decisions we categorize as punitive, such as those affecting arrests, convictions, sentencing, or parole. Any use of an algorithm in the criminal justice system must meet the strictest of legal and ethical standards so as to ensure the rights of individuals affected by such decisions.

Because our program uses an algorithm that will be used only to provide assistance and not to inform punitive decisions, one might suppose that these standards could be somewhat relaxed in our application; however, to the contrary, we have taken the position that we will adhere to the same strict standards that we advise to be used in other contexts.

Thus, to guide the development of our risk model and intervention program, and to educate the public (especially judges, police, and prosecutors), we set out to delineate the issues that must be considered when employing algorithms in the context of law enforcement and the criminal justice system. Toward this goal, we conducted a detailed legal and ethical analysis of U.S. Constitutional principles and of all the appellate court cases (n=101) that reviewed the factors that police officers said they relied on to make predictions about criminality (See Appendix B).

We analyzed the 101 court cases that dealt with law enforcement prediction of criminality and recorded the variables used in the predictions and analyzed how courts determined whether it was legally permissible to use each variable (See Appendix B). We found a total of 125 variables that law enforcement officers said they used to predict that a particular individual was likely to commit a crime or have committed a crime. In the cases we considered, the number of variables that were employed ranged from one to eleven. The court cases, ranging from the trial court level to U.S. Supreme Court, emphasized the importance of context, suggesting that the use of multiple variables at once is the most appropriate approach. For example, while a certain single variable used alone (such as location) is deemed inappropriate by courts, it can be permissible when used with other predictive variables.

We found that courts were more lenient toward law enforcement officials with respect to the factors they used to predict potential dangerousness at airports than elsewhere in public. We determined that the precedents for the public, non-airport settings were more appropriate for our use. Those cases emphasize that it is legally improper to use race as a predictive variable, and we have thus designed our risk models in a way that does not use race as a variable, nor does it use factors that are

linked closely to race, such as place of residence or gang affiliation.

Even beyond its usefulness within this project, our extensive analysis is sufficiently important and unique that we will expand upon it and publish it as an article to guide judges in their work and to guide police departments in their training of officers.

<u>Recommended legal and ethical guidelines:</u>

Based on our analysis of cases and constitutional principles, we fashioned guidelines for the use of predictive algorithms, and we developed our risk model and intervention strategy so as to meet these legal and ethical guidelines. These guidelines are summarized as follows:

Algorithms should not be used by the criminal justice system in punitive ways, such as for sentencing or parole decisions. Crime-risk models should be designed to help people to avoid adverse outcomes, not as a basis to punish them. Constitutional due process requires that decisions that punish an individual should be based instead on analyses that are entirely specific to the individual[3] and not on aggregate estimates of risk in relation to historical data. Algorithms that predict recidivism are inappropriate for sentencing decisions because they focus on the wrong scientific questions. The appropriate question is whether a longer sentence is more likely to deter future crime in a particular individual. That may depend on factors such as mental illness or whether incarceration itself leads to a greater risk of criminality (for example, by providing opportunities for the individual to learn additional criminal skills). Most importantly, the constitutional right of due process requires that the defendant receive an "individualized" sentence and not be sentenced based on aggregate risk predictions. In the 2016 *State v. Loomis* case, the court ruled that "using a risk assessment tool to determine the length and severity of a sentence is a poor fit. As scholars have observed, '[a]ssessing the risk of future crime plays no role in sentencing decisions based solely on backward-looking perceptions of blameworthiness, … is not relevant to deterrence, … and should not be used to sentence offenders to more time than they morally deserve.'"[4] We will be using our risk assessment tool only in instances that offer social services, not for punitive actions such as increasing sentences.

Algorithms must not use variables based on personal attributes, such as race, which a person cannot change. The development of an algorithm can be a way to avoid the idiosyncratic (or even racially-biased) decisions that could be made by police officers in the absence of such a scientific model. Yet care must be taken not to introduce racial biases into the algorithm itself. In Pennsylvania, state police instructed bank employees to take photos of suspicious-looking African Americans— thus setting the stage for creation of a criminal profile that applied only to African Americans. This action was held to create a cause of action under the Civil Rights Act on behalf of an African American man who was photographed.[5] Similarly, courts have held that race cannot be used as a predictive factor in police decisions to stop and frisk an individual. The U.S. Supreme Court has held race to be an impermissible factor when assessing reasonable suspicion,[6] saying that "the Constitution prohibits selective enforcement of the law based on considerations such as race."[7]
In a 2017 U.S. Supreme Court case, *Buck v. Davis*, the defendant's expert argued

that race is a factor "know[n] to predict future dangerousness."[8] The court wrote that "this is a disturbing departure from a basic premise of our criminal justice system: Our law punishes people for what they do, not who they are."[8] We argue that, if carefully designed, a risk model can help to make decision-making fairer, not less fair, by avoiding the idiosyncratic or even racially biased notions that humans may entertain in the absence of a scientific approach based on hard data.

In Year 1, we not only excluded variables relating to race, but we also analyzed the output of our risk model to ensure that racial or ethnic bias did not arise inadvertently, as we explain later.

Algorithms and procedures for their use in crime prevention programs should be publicly disclosed. Numerous for-profit companies offer predictive policing services that involve a proprietary algorithm. If decisions are made about individuals based on a secret algorithm, this will infringe on their Constitutional rights. In 1977, the U.S. Supreme Court held that a defendant's due process rights had been violated when he was sentenced to death based on a report which contained confidential passages that he was not allowed to see or refute.[9]

In an analogous situation, when a company developed a forensic DNA test and refused to disclose its scientific underpinnings to a criminal defendant, a Vermont court prohibited its use on the grounds that the defendant had a right to know how the predictions were made about DNA matching and a right to challenge the methodology.[10] The court said that such secrecy "is more than problematic, it is anti-scientific in that it inhibits the ability of scientists in the field (including defense experts) to test the manufacturer's claims."[10]

In keeping with this recommendation, we are making the full technical details of our model publicly available, and Whitt is actively working to inform the community about the details of the program.

The algorithm must not use variables in ways that infringe rights of free speech, association, and privacy. The most appropriate sources of data for crime-risk modeling are prior arrests and victimization. However, there is an increasing interest among police departments in using a person's social media posts to investigate crimes and predict criminality. Seventy percent of law enforcement agencies use social media posts in criminal investigations.[11] In recent years, several companies sold services to police departments to use social media posts to predict an individual's dangerousness[12] or to monitor activist groups, such as black activists.[13]

The use of a variable based on a social media post in a predictive policing model is scientifically unsound[14] and legally improper. Using a variable based on social media posts in a predictive algorithm could violate a person's right of freedom of expression and freedom of association. It could also violate the rights of the friends and relatives who communicate with him on social media. In an analogous situation, agents of the Federal Bureau of Investigation went to court to ask permission to remotely activate a suspect's webcam to take photos of what the suspect looked like and screenshots of what he was doing on his computer. In that case, the request was turned down because of the intrusiveness of the surveillance and the fact that innocent people who used the same computer (which might have been in a home, a dorm, or even an internet café) might have been photographed as well.[15] Similarly, the American Civil Liberties Union successfully convinced Twitter, Instagram, and Facebook to cease providing data to the companies (Geofeedia, SnapTrends, and

Media Sonar) offering predictive services based on social media posts to police departments.[13]

We have carefully avoided the use of any kind of information that goes beyond standard crime records so as to avoid the pitfall illustrated by the aforementioned cases.

Technology research: Development of the crime risk model:
We obtained from the EPD a crime database using Microsoft SQL Server. This database contains a total of 6,828 crime incidents during the period from 01-Jan-2012 to 25-Jul-2016, involving a total of 22,049 individuals. For privacy protection, we asked the EPD not to provide us with any sensitive details about these individuals, such as name, sex, race, and home address, so as to exclude them from our model development. We later requested race data from the EPD for the sole purpose of testing to ensure that no racial bias inadvertently crept into the model's risk assessments (the model passed this test, as we explain later).

In the database, the crime incidents are stored in various data tables, including general crimes, narcotic arrests, warrants, gang affiliations, bookings, etc. As explained earlier, the EPD identified a set of "serious" crimes for developing our risk model of an individual becoming an offender or victim. These identified crime types are listed in Table 1 along with their Illinois Uniform Criminal Record (IUCR) codes.

In our risk model, we assigned a risk score only to those who have been arrested at least once in the past two years. The development of our risk model involves training and testing phases. To establish the platform for this process, we used the crime data from 01-Apr-2012 to 07-Jul-2016 (four years and three months). We created two datasets, each of which consists of two years of data plus a one-year look-ahead period to serve as example data from which the machine-learning risk model could learn to predict risk. The two datasets consist of the following: 1) Dataset 1: 11,165 individuals among which 1,101 were involved in a serious crime; and 2) Dataset 2: 10,738 individuals among which 1,130 were involved in a serious crime. Using these datasets, we identified 24 potentially predictive variables that we have used in the predictive algorithms. [16]

Model versions:
In Year 1, we undertook an iterative process for development of the risk model through the application of machine learning technology. *Machine learning* refers to a mathematical approach in which a computer algorithm makes assessments based on data that are provided to it, with the unique aspect being that the algorithm *learns* to make these assessments based on historical data for which the correct answer is known. For example, an algorithm can be trained to evaluate an individual's risk for future involvement in crime by presenting it with historical data about individuals in the past, including the records leading up to a given moment in time, and the actual outcome that ensued in a subsequent period. As in any engineering design project, machine learning often involves an iterative process in which various concepts are studied, compared, and refined. During Year 1 we developed candidate model versions, identified as v0.x, among which we selected v0.3 as the best model to date, and designated it as v1.0, meaning that it is the first version to be released to the EPD.

Risk model v0.1 Model v0.1 was an implementation of the starting-point concept that we had described in our initial proposal. Because Elgin is focused on a broad array

of serious crimes (not only shootings), we sought to predict an individual's risk for involvement in at least one serious crime incident in the upcoming year. Model v0.1 proved to be highly predictive of involvement in crime; however, we realized that it was too generic, in that it made no effort to differentiate criminal involvement in terms of extent or severity of the crimes. Suppose, for example, that Person A will be involved in a simple assault next year, while Person B will be involved in one shooting, two aggravated battery incidents, and an armed robbery. In this case, v0.1 would classify both persons identically because both were involved in at least one serious crime. But this would fail to capture the obvious conclusion that Person B is likely to experience a much greater extent and severity of crimes than Person A.

Crime severity value	Code	Description
100	110	First Degree Murder
100	115	Homicide of an unborn child
60	280	Predatory criminal sexual assault of a child
30	170	Drug induced homicide
30	190	Solicitation for murder or murder for hire
30	261	Aggravated criminal sexual assault
30	281	Criminal sexual assault with an object
30	310	Armed robbery
30	326	Aggravated vehicular hijacking
30	480	Heinous battery
30	485	Aggravated battery of a child
20	262	Forcible sodomy
15	120	Voluntary manslaughter of an unborn child
15	130	Second Degree murder
15	260	Criminal sexual assault
15	325	Vehicular hijacking
15	330	Aggravated robbery
7	320	Robbery
7	487	Aggravated battery of an unborn child
7	495	Aggravated battery of a senior citizen
5	141	Involuntary manslaughter
5	142	Reckless homicide
5	160	Concealment of a homicide death
5	165	Involuntary manslaughter and Reckless homicide of an unborn child
5	410	Aggravated battery
5	488	Aggravated domestic battery
5	491	Aggravated stalking
1	460	Battery
1	470	Reckless conduct
1	475	Battery of an unborn child
1	486	Domestic battery
1	510	Aggravated assault
0.1	560	Assault
0	150	Justifiable homicide

Table 3. Crime severity values used in v0.3 and v0.4 based on maximum sentences under sentencing guidelines in the State of Illinois.

Risk model v0.2

To address this shortcoming, we developed model v0.2, in which we shifted to predicting an overall *crime-risk index* (CRI) for each individual, based on both the extent and severity of crime involvement. To capture the notion of severity, we adopted a numeric scale, in which each type of crime is assigned a *crime severity value* (CSV). An individual's CRI is then computed as a count of crime incidents, with each incident weighted by its CSV. For example, suppose that homicides are assigned a CSV of 100, while the less-severe crime of robbery is represented by a CSV of 25. In this case, an individual involved in one homicide and two robberies would have a CRI of 1×100 + 2×25 = 150.

As a proof of concept for our new modeling approach, we assigned a coarse,

preliminary system of CSVs for various categories of crime (See Table 2). We trained a model to predict an individual's CRI for the upcoming year, based on crime involvement in the past two years.

We trained and tested numerous potential models based on the 24 variables that we found to be predictive. The best-performing model proved to be one known as the least absolute shrinkage and selection operator (LASSO), which not only produced the best performance metric, the normalized discounted cumulative gain (NDCG), but also required only eight of the variables to be effective, thus yielding a readily interpretable model.

Risk model v0.3

Having achieved good results with the concept of the crime severity index (CRI), we decided to replace the somewhat arbitrary weights that we assigned in v0.2 to various crime types. In model version v0.3, we modified the weighting scheme to define crime severity by using the maximum sentence (in years) prescribed by law in the State of Illinois (See Table 3). The reasoning behind this approach is that society has already quantitatively defined the relative severity of crimes within the sentencing guidelines, thereby giving a sound foundation for our risk assessments. This resulted in an algorithm that provides even better results than v0.2. Version 0.3 (v0.3) only required seven variables to obtain optimal performance.

Risk model v0.4 Version 0.3 (v0.3) made use of a variable called "felony," which encompassed multiple subtypes of crime. Normally, one would avoid using variables that contain redundant information (a murder is also a felony); therefore, in v0.4, we dropped felony as a separate variable. Interestingly, while this improved the model accuracy for individuals exhibiting relatively lower risk levels, it slightly reduced performance for individuals in the very high risk category that is the focus of the project. After significant discussion with the EPD, we concluded v0.3 is preferable to v0.4. Since v0.3 is better than expected, and certainly more than adequate to use as the basis for the pilot phase of the social services intervention, we designated it as v1.0 (i.e., the version that will be used in the initial interventions).

Evaluation of risk model v1.0:

Simplicity of the model It is generally desirable for a risk model to use a small number of variables so that the end user can readily understand how it works and what it means. Model v1.0 reduced the 24 variables that we considered to a shorter list of just seven variables that the model found to be most important for evaluating risk, with importance represented by the coefficient used by the LASSO model to weight them (See Figure 1).*Accuracy of the model*
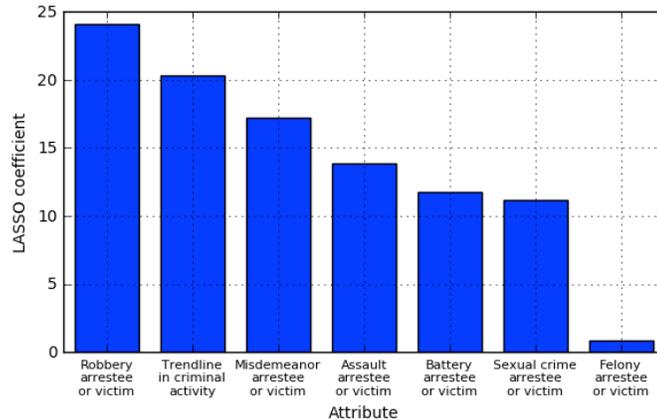

Figure 1. Variables used by model v1.0 in decreasing order of importance for determining risk, as expressed by the weighting coefficient used by the LASSO. All of the variables are based on involvement in crime during the past two years. The "Trendline" expresses the rate at which the individual's involvement has been increasing/decreasing during the same period.

Because the goal of our risk model is to rank individuals based on crime risk, the problem is mathematically equivalent to that of a web search, in which the search engine attempts to rank web pages in order of relevance to the user. Thus, we have adopted similar metrics to evaluate our risk model and to guide its training.

The performance metric that we used as the parameter to optimize when designing the risk model is the so-called *normalized discounted cumulative gain* (NDCG), a metric to test the accuracy of a ranking scheme. In web search, the NDCG focuses on correct ranking of the highest-ranked web pages (for instance, the ones that will appear first in a user's Google search); in our application, the NDCG focuses on correctly ranking the highest-risk individuals (those that will be invited to participate in the social service interventions).
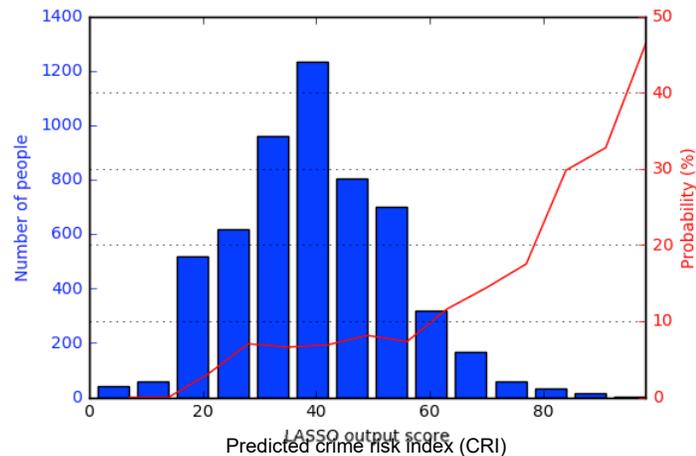

Figure 2. Histogram (blue bars) showing the number of individuals having various degrees of predicted risk and actual probability that these individuals will become involved in a serious crime (red curve). Among the 91 persons with the highest level of risk, the red curve shows that roughly 46% of them will actually be involved in a serious crime in the upcoming year. This demonstrates that the model is correctly identifying the most at-risk people and that the number of such people is small enough that the social work team can make a useful impact.

Although the NDCG is an appropriate metric for quantitative analysis, it is not terribly intuitive. Therefore, to understand the results, we instead look at the probability that individuals judged to be at the highest levels of risk will actually become involved in a serious crime in the upcoming year. Figure 2 shows that individuals with the highest predicted crime risk (near 100), which are few in number, have a 46% probability of being involved in a serious crime. This shows that the risk model is correctly identifying the most at-risk people and clearly differentiating them from the vast majority, whose risk ranges from 0-10%.

A different view of the data is seen in Figure 3, which shows cumulative crime-risk index (CRI) for the individuals with the highest risk. The graph shows that 200 individuals account for 30% of the total crime severity. These are just 6% of the individuals arrested in the past two years and amount to a mere 0.18% of the population of Elgin. Thus, reaching out to this group with support can have a significant impact on crime.
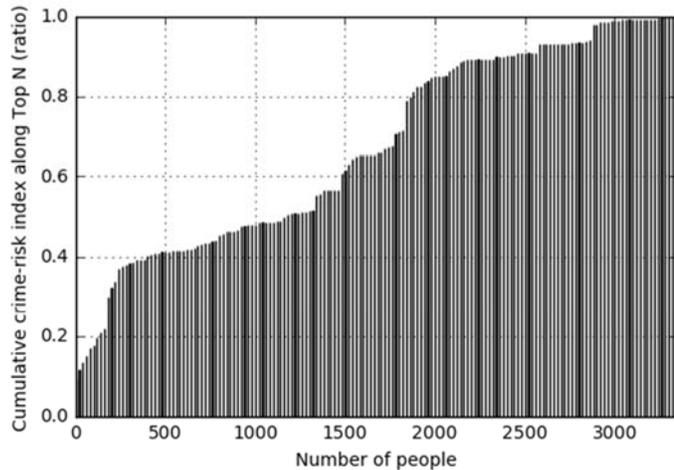

Figure 3. Normalized cumulative crime-risk index (CRI), showing that the 200 individuals with the highest predicted risk account for 30% of the total CRI. In other words, 6% of recent arrestees account for 30% of overall crime severity, and these individuals constitute just 0.18% of Elgin's population.

Ruling out racial and ethnic bias Our model does not make use of race or ethnicity, nor does it use variables from which these factors may be readily inferred, such as gang membership. Nevertheless, we felt it important to perform a statistical test to rule out the presence of bias in the risk assessments.

We define model bias as follows. Suppose that individuals in a population belong to either racial/ethnic group A or B. A risk model is biased against group B if, among persons in groups A and B who are *predicted* to be at high and equal risk, the *actual* risk levels for group B are systematically lower than for group A. In other words, if we were to declare persons belonging to group B to be at high risk when actually they are not, then the risk model would be biased against group B.

To test our model, we looked at actual risks for individuals having the top 500 risk scores for three different time periods and computed the group means and standard deviations of the actual risks. We divided these individuals into three groups: white, Hispanic, and African American. (We omitted Asian/Pacific Islanders and American Indian/Alaskan Natives because the datasets contained too few individuals from these groups to permit analysis.) For the persons with the top 500 risk scores, we found the mean actual risk was 23% for whites and Hispanics and 25% for African Americans. This would imply a slight bias *in favor of* African Americans, but the result

was not statistically significant under any standard definition of significance (the p-values for all pairwise comparisons ranged from 0.40 to 0.74). Social Services Intervention Appendix A is a draft of the official Standard Operating Procedure (SOP) for our social services intervention program, which is titled "Social Services Crime Prevention Program." This SOP document was developed through extensive collaboration between us and the EPD. The key aspects of the SOP are as follows:

The program will be based on social services only, and no law enforcement action will be taken against any individuals identified as at-risk by our model. To be sure this is clear to the participants, they will be invited to participate by either the head social worker (Stingley) or the community liaison (Whitt). Meetings with participants will take place either at their homes by appointment or in the social services area of police headquarters.

The intervention program will be overseen by the head social worker (Stingley). She will follow up on each individual's case to ensure that good progress is being made, involving regular follow-up calls or visits with each participant.

Potential participants will be selected based on recommendations of the risk model, with final decision made by a committee including diverse members of the EPD. For example, the EPD may know that an individual is deceased, incarcerated, or no longer a resident of Elgin. The EPD may also know that the person has been consistently unwilling to participate in past social services programs. The draft SOP document (Appendix A) is currently undergoing approval by Elgin. As soon as the SOP is finalized, the committee will convene to review the recommended candidates and the interventions will begin shortly thereafter.

Plans for Year 2:

Monitoring and evaluating the interventions

In Year 2 we will employ a process to gather data about the interventions so that we can evaluate their effectiveness and propose improvements where appropriate. We will develop a protocol for the committee that selects the participants, including a record of reasons given for exclusion of any participant. We will gather anecdotal data on the progress of the participants so as to provide a qualitative assessment of the program to obtain lessons for future similar efforts. We will attempt to evaluate success quantitatively as well, but we do not anticipate being able to demonstrate statistical significance during the short span of the project and the relatively small number of participants. To perform statistical comparisons, it would not be appropriate in this context to randomly assign individuals to treatment and control groups as one would do in a classic randomized controlled trial. However, we will be able to use techniques from quasi-experimental design, such as comparing the high-risk individuals who participate in the program to those not yet invited and by comparing to those who declined to participate. According to the SOP, the social workers will be providing regular updates on the progress of individual participants, and these will be reported to the program administrator on a quarterly basis.

Improved risk models

As we did in Chicago's program based on risk models, the Elgin risk model will undergo a process of continual improvement. In Year 2, we will study the use of information

about co-arrestees or co-victims to enhance the risk assessments. We will also incorporate various mechanisms to make the risk assessments more time-sensitive, based on our prior experience that very recent crime incidents are much more significant to an individual's risk than incidents that took place weeks or months ago. We will investigate whether sufficient time-sensitivity can be gained to allow interventions to be made on an urgent basis. For example, if an individual's close associate (associated by co-arrests or co-victimization) has just been the victim of serious crime, it may be beneficial to reach out to that individual to identify the circumstances and determine whether intervention can help.

Influencing the national debate / publicizing our work

We anticipate that a central focus of Year 2 will be to educate judges, police, and the scientific and legal communities about the concepts behind this project. We will set forth our proposed legal and ethical guidelines for the use of crime risk models, conduct media interviews about the program, and publish scientific papers on the technology involved. This project will draw attention to Illinois Institute of Technology, highlighting our ability to translate technological breakthroughs into solutions to challenging societal issues of the day.

Attachments:

Appendix A: Standard operating procedure (SOP).

	ELGIN POLICE DEPARTMENT 151 Douglas Avenue Elgin, Illinois 60120	
Effective Date: (insert later)	**STANDARD OPERATING PROCEDURE**	**Revised Date:**
Chief of Police:	Social Service Crime Prevention Program, 42.9	
Cross Reference:	**Policy Sections:** 42.9.1 Program Administration 42.9.2 Program Selection 42.9.3 Department Review & Eligibility for Services 42.9.4 Intervention 42.9.5 Program Evaluation	

PURPOSE

The purpose of this policy is to establish guidelines for the administration of the department's Social Service Crime Prevention Program (SSCPP).

POLICY

It is the policy of the Elgin Police Department to endeavor to create a safer community. The SSCPP is a collaboration with Illinois Institute of Technology (IIT) to identify individuals with a high propensity for being a party to a serious crime in the capacity of a victim or offender. The department, in conjunction with participating social service agencies, aims to provide the appropriate services to program participants to prevent his/her involvement with a future crime.

DEFINITIONS

Risk model: A mathematical procedure based on statistics that estimates the chances of something happening in the future based on risk factors. For example, smoking is a risk factor for cancer, because it increases the chances. Similarly, a consistent pattern of involvement in crime incidents is a risk factor for a person's continued involvement.

PROCEDURES

42.9.1 PROGRAM ADMINISTRATION

A.   The program will be administered by a supervisor designated by the chief.

B.   The head department social worker will oversee the intervention portion of the program.

C.   The head social worker will report on the status of the intervention on each person in the program to the program administrator at least bi-weekly.

D.   The program administrator will track all persons in the program and determine when additional people shall be added. A person will generally be added when a person currently in the program has completed services, refuses initial services, or ends participation.

E.   A department information technology specialist will work with the program administrator and IIT on any technical needs for the program.

42.9.2 PROGRAM SELECTION

A.   The department will provide data to IIT personnel for the exclusive purpose of creating a risk model that will help guide the selection of participants who might form the SSP. The data that will be provided to IIT is confidential, and IIT personnel shall not release or utilize the data for any other purpose.

B.   The data may include, but are not limited to arrests/bookings, incidents, criminal reports, and parole status. The data provided to IIT will not include a person's name, race, ethnicity, physical characteristics, or place of residence to avoid potential of bias in the results (except to the extent that racial data may be used apart from the risk analysis to double-check that the algorithm has no racial bias). All persons provided in the data set will be identified by their unique jacket number, as listed in the department's records management system.

C.   IIT will develop a risk model that will identify individuals having a high risk of being involved in a serious crime so that the social service interventions can be focused on individuals who most need it.

D.   The results of IIT's risk analysis will be forwarded to the program administrator.

E.   The analysis will contain the following information:
   1. An estimated risk level for each individual with a recent arrest history
   2. An explanation of how each individual's risk factors led to the estimated risk
   3. The relative severity (compared to other individuals) of each of the

individual's risk factors, expressed as a percentile

42.9.3 DEPARTMENT REVIEW AND ELIGIBILITY FOR SERVICES

A. A review of the data set will be conducted by a committee composed of diverse members of the department with varying job responsibilities. The review process will be coordinated by the program administrator. The purpose of this review is to ensure that the individuals identified in the analysis are appropriate candidates for the program. The review shall consider, but not be limited to, the following information:

1. Current custody status with the Illinois Department of Corrections (IDOC)
2. Residency: The individual resides in the city limits.
3. Pending or recent police contact, to include types of contact
4. Likelihood of participating in the program

B. Records shall be maintained for those individuals not selected; these records shall provide a brief explanation as to why the person was not selected to participate.

C. Individuals under the age of 18 are not eligible for this program. Furthermore, data pertaining to juveniles is excluded from the data provided to IIT.

42.9.4 INTERVENTION

A. Initial contact with selected individuals will generally be conducted by a sworn officer and a social worker or department community outreach specialist. When possible, the initial contact will be made at the individual's home, by appointment set up by the social worker. If the individual wishes to meet at the station, the meeting should be conducted in the social services area or a similar type of space.

B. During the initial contact, the individual should be given information about the program and the desire for the department to facilitate services to reduce the likelihood of involvement in criminal activity. An appointment should then be scheduled for the subject to meet with a department social worker for an initial evaluation for services.

C. The head social worker shall assign a social worker to work with the subject. The program administrator will also assign an officer who will serve as a liaison to the individual.

D. During the initial evaluation, the assigned social worker and the individual shall determine which of the available services the individual might need to help prevent future involvement in a serious crime, either as a victim or as a perpetrator. These services may include education, job training, mental health services, and substance abuse treatment. Referrals shall then be made to the appropriate partner agencies.

E.   Following the referrals, the social worker and liaison officer shall make regular follow-up calls or visits with the individual to track progress and assist with services.

F.   The liaison officer shall also make a designation in the department records management system to track any police contacts with the subject.

42.9.5 PROGRAM EVALUATION

A.   Quarterly, the program administrator will provide to the deputy chief a report showing the progress of each individual selected in the program. The report will include progress on social service utilization and any criminal activity involvement by the individual.

B.   The completed report will be provided to IIT, which will conduct an evaluation of the

rogram to determine its effectiveness and to provide further insight on any modifications and suggested improvements.

Appendix B: Spreadsheet summary of case law, attached as an Excel spreadsheet.

1. J. Walker. "Roskam, Elgin police honor department's 'unsung hero.'" *Chicago Tribune.* (March 9, 2016).

2. J. Walker. "Personal, institutional racism focus of Black Lives Matter discussion." *Chicago Tribune.* (February 26, 2016).

3. *Wisconsin v. Gallion.* 678 N.W.2d 197, 209 (Wis. 2004).

4. *State v. Loomis.* 881 N.W.2d 749, 770 (S. Ct. Wis. 2016).

5. *Hall v. Pennsylvania State Police.* 570 F.2d 86 (3d Cir. 1978).

6. *Whren v. United States.* 517 U.S. 806, 810 (1996).

7. *Whren v. United States.* 517 U.S. 806, 813 (1996).

8. *Buck v. Davis.* 137 S.Ct. 759 (2017).

9. *Gardner v. Florida.* 430 U.S. 349, 352 (1977).

10. *State v. Pfenning.* No. 57-4-96 Grand Isle County, Vermont (Apr. 6, 2000) (unpublished decision).

11. "2016 Law enforcement use of social media survey." *International Association of Chiefs of Police* (February 2017).

12. K. Leetaru. "Geofeedia is just the tip of the iceberg: the era of social surveillance." *Forbes* (October 12, 2016).

13. M. Cagle. "Facebook, Instagram, and Twitter provided data access for a surveillance product marketed to target activists of color." *American Civil Liberties Union Blog* (October 11, 2016).

14. M. L. Williams, A. Edwards, W. Housley, P. Burnap, O. Rana, N. Avis, J. Morgan, L. Sloan. "Policing cyber-neighbourhoods: tension monitoring and social media networks." *Policing and Society* **23**, 1-21 (2013).

15. *In re Warrant to Search a Target Computer at premises Unknown*, 958 F.Supp.2d 753 (S.D. Tex. 2013).

16. The variables are: Frequency of arrests or victimization; Individual's age during the latest arrest; Trend in criminal activity; Homicide arrests; Arrestee or victim of sexual assaults; Arrestee or victim of robberies; Arrestee or victim of batteries; Arrestee or victim of domestic violence; Bookings; Narcotic arrests; UUW arrests; Arrestee or victim of home invasion; Juvenile arrests; Arrestee or victim of burglaries; Shooting arrests; Victim of shooting incident; Arrestee or victim of violent crime; Felony arrests; Local offender; Arrestee or victim of misdemeanor; Number of first-degree connections in the social network to individuals associated with serious crimes; Number of second-degree connections in the social network to individuals associated with serious crimes; Member of the largest cluster in the social network; Arrestee or victim of assaults.