The New York Times

DANGER ZONE

Inside the Algorithm That Tries to Predict Gun Violence in Chicago

By Jeff Asher and Rob Arthur

June 13, 2017

Gun violence in Chicago has surged since late 2015, and much of the news media attention on how the city plans to address this problem has focused on the Strategic Subject List, or S.S.L.

The list is made by an algorithm that tries to predict who is most likely to be involved in a shooting, either as perpetrator or victim. The algorithm is not public, but the city has now placed a version of the list — without names — online through its open data portal, making it possible for the first time to see how Chicago evaluates risk.

We analyzed that information and found that the assigned risk scores — and what characteristics go into them — are sometimes at odds with the Chicago Police Department's public statements and cut against some common perceptions.

■ Violence in the city is less concentrated at the top — among a group of about 1,400 people with the highest risk scores — than some public comments from the Chicago police have suggested.

■ Gangs are often blamed for the devastating increase in gun violence in Chicago, but gang membership had a small predictive effect and is being dropped from the most recent version of the algorithm.

■ Being a victim of a shooting or an assault is far more predictive of future gun violence than being arrested on charges of domestic violence or weapons possession.

■ The algorithm has been used in Chicago for several years, and its effectiveness is far from clear. Chicago accounted for a large share of the increase in urban murders last year.

What Is the Strategic Subject List?

The Strategic Subject List was created by the Illinois Institute of Technology in 2013, according to the university's Miles Wernick, the lead researcher on the project, and stays away from variables that could discriminate, like race, gender and geography.

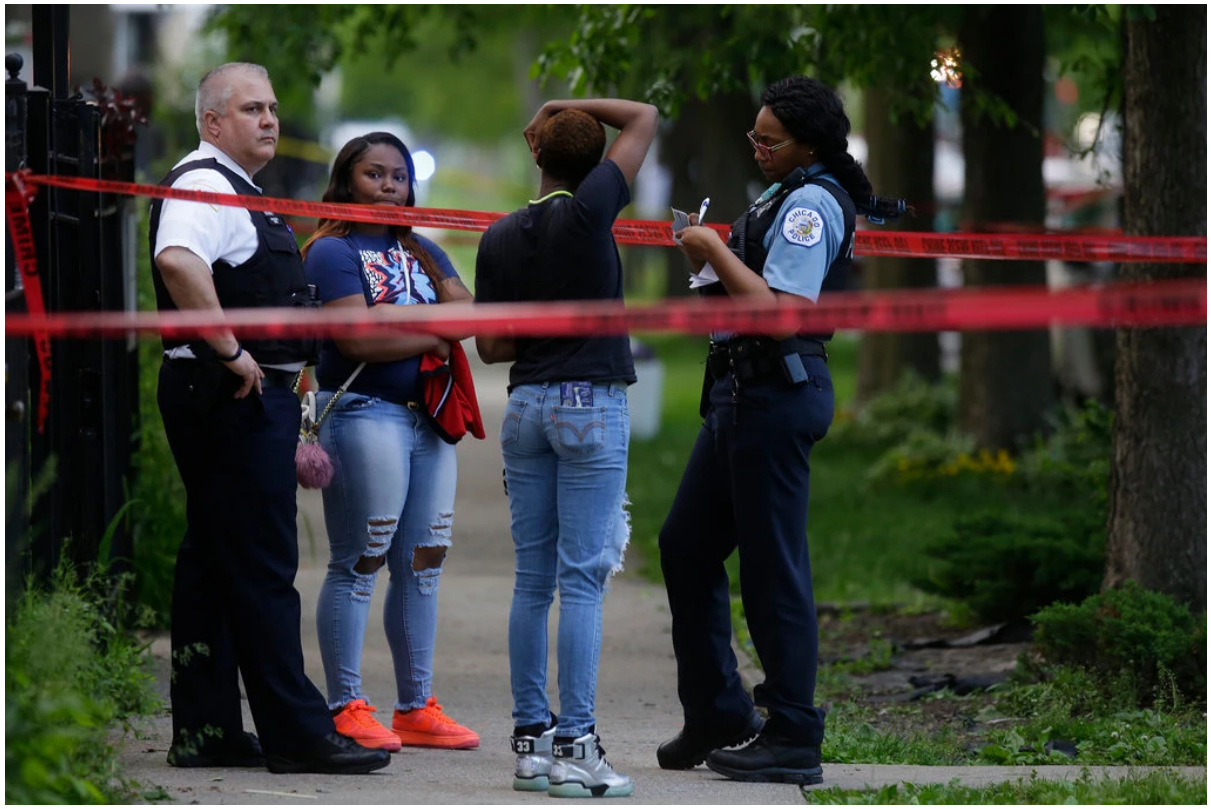
The algorithm is applied to hundreds of thousands of arrested subjects, and helps prioritize limited resources to focus on those at highest risk. It does not create separate risk scores for who will shoot or be shot; the research is supposed to focus on victimization because estimating a person's chances of being a perpetrator creeps eerily close to the movie "Minority Report."

But some have accused the city of using the risk scores as a secret system to punish citizens for crimes they haven't yet committed. The risk scores have been used numerous times to guide large-scale roundups, in addition to individual meetings to warn those at high risk of the danger they're in and to offer them services to try to turn their lives around.

Similar experiments are taking place in dozens of police departments across the country. Civil liberties groups question the legality and efficacy of such programs, saying citizens, besides having the right to a presumption of innocence, should be able to know exactly how they wind up on a list, and should be able to challenge their inclusion.

Risk scores in Chicago range from 0 to 500, with higher scores representing increasing risk. The publicly available data set is huge, with almost 400,000 people, although the vast majority of them have relatively low risk scores.

"What we know from past criminological/social science research is that a small percentage of the population drives violence," said Andrew Papachristos, a professor at Yale whose research into gun violence within social networks inspired creation of the algorithm. "Small numbers of individuals or groups or street blocks are responsible for a disproportionate amount of violence. That's the basics for why something like this may be useful."



Police officers talked to two women as they investigated a fatal shooting that killed a 15-year-old boy in Chicago last month. The city has experienced a devastating increase in gun violence in recent years. Joshua Lott for The New York Times

Mr. Papachristos has distanced himself from the Strategic Subject List, expressing concerns about its transparency and about its potential use to forecast who will become shooters, according to a recent report in The Sun-Times.

To date, the Chicago Police Department has declined to release details of the algorithm, citing proprietary technology. (Last week, The Chicago Sun-Times and three independent journalists filed a Freedom of Information Act suit against Chicago and its police department to release full information on the algorithm.)

But using the publicly available data that they have released, we reverse-engineered the impact of each characteristic on the final risk scores with a linear regression model. Because the department didn't release all the information that the algorithm uses, our estimates of the significance of each characteristic are only approximate. But using what was available to us, we could predict

risk score very accurately, suggesting that we are capturing much of the important information that goes into the algorithm.

The most significant characteristic for computing an S.S.L. risk score is the age of a potential victim or offender. For every decade of age, the risk score declined by about 40 points. Practically speaking, this variable limits the list to young people: No one older than 30 falls within the highest-risk category with a score at or above 480.

The other most important risk factors centered on criminal activity. Although arrest-related variables did play a significant role, a far more crucial one was whether a person had been the victim of a crime.

In particular, victims of assault and battery or shootings were much more likely to be involved in future shootings. Arrests for domestic violence, weapons or drugs were much less predictive. Gang affiliation, which applied to 16.3 percent of people on the list, had barely any impact on the risk score.

The algorithm has been updated several times, and Mr. Wernick noted that the variables of gang affiliation and narcotics arrests were dropped from the most recent version. Those with a gang affiliation make up roughly 2.4 percent of Chicago's population of 2.7 million — so many people (nearly 65,000) that it may explain why it has little predictive value.

How the Results Differ From Police Statements

Although we can work backward to determine what factors affect S.S.L. scores, further analysis is challenging because scores are regularly adjusted. For example, if a person with a risk score of 460 was shot in April 2016, his score will probably be around 20 points higher in July 2016. The effect of these changes is relatively small, though, and does not change the overall assessment of the risk scores' predictive capabilities.

About 2,250 people on the list were "Party to Violence" — meaning they were involved in a shooting or a murder — most recently in 2016. The higher the score, the more likely a person was to be involved in a shooting or a murder, as shown in the table.

Over a third of those with a risk score of 500 were involved in a shooting or a murder in 2016. Mr. Wernick said about "one-third of the top 400 will be involved in a shooting in the next 18 months." The data appear to back up that assertion.

Although the algorithm evaluates risk well, the gun violence at the highest levels of risk is not as concentrated as some Chicago Police Department comments have suggested. A New York Times article in May 2016 quoted the police chief, Eddie Johnson, as saying that "about 1,400 are responsible for much of the violence."

A CBS Chicago report from May 2016 cited the police department as saying that 1,500 people from the S.S.L. are responsible for the majority of gun violence in Chicago.

The top 1,400 people have scores of 429 and up, and they are in fact disproportionately involved in violence. But that disproportionate share amounted to less than 20 percent of the total gun violence in Chicago in 2016.

The Chicago Police Department declined to comment on the accuracy of Chief Johnson's statement.

The risk scores may be useful in predicting violence, but their effectiveness as part of Chicago's crime-fighting arsenal is in question. Murder in Chicago remains considerably higher than it was before violence started rising two years ago.

Maybe crime would be even worse without the help of risk scores; it's hard to know, because multiple factors can cause fluctuations in crime. But it seems clear that maximizing the power of the algorithm's scores may require identifying new ways of preventing risk from becoming reality.

Jeff Asher is a crime analyst who runs the NOLA Crime News data analysis blog. You can follow him on Twitter at @Crimealytics. Rob Arthur is a freelance journalist writing about crime, politics and baseball. You can follow him on Twitter at @No_Little_Plans.