Document Title: Using Machine Learning to Identify High Risk Domestic Violence Offenders in NYC, Final Summary Overview

Author(s): Jens Ludwig

Document Number: 305443

Date Received: October 2022

Award Number: 2017-VA-CX-0033

Using Machine Learning to Identify
High Risk Domestic Violence Offenders in NYC

Final Summary Overview

Department of Justice, Office of Research and Evaluation, National Institute of Justice

• Grant award #	2017-VA-CX-0033
• Award amount	$452,553
• Grantee Organization	University of Chicago
• Grant Start Date and End Date	January 1, 2018 – April 30, 2022
• Project Title	Using Machine Learning to Identify High Risk Domestic Violence Offenders in NYC
• Principal Investigator (PI)	Jens Ludwig
• PI e-mail address	jludwig@uchicago.edu
• PI phone number	773-702-3242

I. Project Summary

a. Background

Over the past twenty-five years, homicides in New York City have plummeted nearly 85%. While debate continues about the causes of this decline, most observers believe a role was played by the New York City Police Department (NYPD)'s "CompStat" system, which capitalizes on growing availability of "big data" to help predict trends and focus resources on crime "hot spots." Still, one major source of homicide in New York City has proven mostly resistant to this decline: domestic violence (DV). Over this period, the share of New York City's DV-related homicides has more than tripled. Victim-centered approaches to reducing serious victimization are promising, but also expensive, making it critical to effectively target services to those who need it the most.

Why has domestic violence followed such a different trajectory from that of other forms of inter-personal violence in New York City (and nationwide)? We believe a key challenge is the much greater difficulty (relative to "street crime") in predicting when and where domestic violence will occur, which in turn makes it difficult to effectively target scarce criminal justice resources where they will be most helpful. Domestic violence is difficult to predict partly because victim reporting rates seem to be very low: estimates range from 1 in 4 to as low as 1 in 50.[1,2] Because so much domestic violence occurs indoors, these events are difficult for police to learn about absent a victim report. Efforts to effectively use data to target law enforcement and other prevention services are

[1] Klein, A. R. (2009). Practical implications of current domestic violence research: For law enforcement, prosecutors and judges.
[2] Tjaden, P., & Thoennes, N. (2000). Prevalence and consequences of male-to-female and female-to-male intimate partner violence as measured by the National Violence Against Women Survey.

challenging when the available data are so limited and their "signal" so difficult to extract. This makes it crucial to use tools that extract as much "signal" as possible from available data.

The purpose of this project—a collaboration between the University of Chicago Crime Lab and NYPD—was to address this "data gap" by developing and testing a novel machine learning-based statistical model to predict the risk of domestic violence victimization in order to improve targeting of domestic violence resources in New York City. We initially aimed to conduct this research in two phases. In the first phase, we developed a state-of-the-art statistical model using machine learning techniques to predict victimization risk among DV victims in New York City. To build our tool, we assembled several million records of NYPD administrative data covering all of New York City. In phase two, we formally tested the ability of our machine learning (ML) tool to better identify those at risk of repeat domestic violence through a large-scale randomized controlled trial (RCT). The RCT would allow us to identify which people—those selected by the ML tool or those selected by NYPD officers—were at higher risk of revictimization. It would also allow us to measure the effects of targeting a promising intervention—home visits conducted by domestic violence officers (DVOs)—using machine learning predictions relative to status quo officer selections. However, we were not able to successfully complete the RCT, and thus, added an additional third phase which would employ a quasi-experimental research design in order to assess the differences in the selections of the algorithms and DVOs. However, due to the Covid-19 pandemic, the work on the project has been put on pause; however, we intend to complete it after the conclusion of this grant.

b. Major Goals and Objectives

In this project, we aimed to

1) develop a novel machine-learning based statistical model to predict the risk of domestic violence re-victimization in order to improve targeting of domestic violence;
2) carry out a randomized controlled trial (RCT) in partnership with NYPD to test the ability of this new tool to reduce repeat domestic violence victimization in the field; and
3) disseminate the findings and larger intervention strategy nationwide.

c. Research Questions

We sought to answer the following questions:

1) Who picks riskier victims, our statistical model or NYPD's DVOs?
2) What is the treatment effect of DVO home visits on violent felony DV revictimization?

d. Research design, methods, analytical and data analysis techniques

Below, we describe our planned research activities in detail. First, we discuss the development of our statistical model, include a description of the types of administrative data used, the construction of the ML model, and the text analysis from Phase 1. We next describe the field intervention and randomized controlled trial from Phase 2.

PHASE 1: DEVELOPMENT OF OUT MACHINE LEARNING STATISTICAL MODEL

Our machine learning tool incorporates NYPD administrative datasets covering all of New York City for the period between January 2006 and January 2017, including: 1) domestic incident reports (DIRs); 2) criminal complaints; 3) arrests; and 4) aided reports. From these four sources, we aimed constructed a victim-level dataset that describes an individual's history of law enforcement contact. There is, however, no unique identifier for victims in NYPD data (and most other jurisdictions), thus the research team developed a probabilistic record linkage algorithm to identify which records belonged to the same person within and across data sources.

To generate predictions of violent felony DV victimization over a 12-month follow-up period, we used eight years of pre-period data (2006-2014) to predict for outcomes in a one-year follow-up period. In order to get a sense of how our model would perform out-of-sample, we divided our data into a "training set" that we used to train our algorithm on an out-of-sample "test set." In the end, we chose a stacking algorithm which combined the outputs of tree-based and linear models. In order to formally test our ML tool's performance with novel data (i.e., in the real world) we designed a randomized controlled trial, which we launched during Phase 2.

PHASE 2: FIELD INTERVENTION AND RANDOMIZED CONTROLLED TRIAL

Our field intervention with NYPD launched in July 2017. Each NYPD command maintains a list of high-priority individuals who are thought to be at risk of serious domestic assault. Individuals placed on this list receive regular home visits from one of the local NYPD command's domestic violence officers (DVOs) in order to reduce their risk of future victimizations. Our field intervention involved 60 NYPD commands: 30 commands were randomized into treatment and another 30 were randomized into control. In treatment commands we added two additional people per DVO to the command's high-priority list. One individual was selected by our algorithm, and one was selected by the command's DVOs. These individuals received NYPD's standard intervention of home visits by DVOs. Control commands did not receive additional victims (i.e., "business-as-usual").

We sought to answer the following questions:

1. Who picks riskier victims, our statistical model or NYPD's DVOs?
1. What is the treatment effect of DVO home visits on violent felony DV revictimization?

In order to measure <u>baseline risk</u>, we compared our tool's picks to those of NYPD DVOs in treatment commands. Conversely, to measure the treatment effects of the home visit intervention, we compared our ML tool's picks in treatment commands to our ML tool's picks in control commands.

Our RCT proved difficult to execute in the real world, where the needs of the research can often conflict with a program partner's operational constraints. These constraints greatly handicapped our model's performance and the power of the RCT. We will discuss these limitations in greater detail in the next section.

e. Expected applicability of the research

We believe that upon the completion and dissemination of our work, our results will be relevant to researchers and policymakers who are assessing the value of statistical decision aids in reducing instances of domestic violence. In particular, we expect our research to shed light on the utility of risk assessments that are built using police data as compared with victim reports, as well as the differences in the types of high-risk cases that are identified by officers as opposed to algorithms.

II. Participants and other collaborating organizations

University of Chicago Crime Lab

The New York City Police Department

III. Changes in Approach to Design

Our RCT proved difficult to execute in the real world as a number of complications arose which affected both our study's statistical power as well as our tool's performance. We highlight some of these challenges below.

First, due to operational challenges and shifting NYPD priorities, our field experiment ran for only seven out of the twelve months intended, severely limiting the number of treated participants and, thus, our statistical power. Second, our risk predictions were, out of necessity, performed with stale data: when our study launched in July 2017, our NYPD administrative data only ran through the end of 2016, severely handicapping our model's performance. Additionally, because our data were relatively old, victim address information was also quite old, leading to an extremely high "vetting" rate of victims. That is, it turned out to be extremely difficult for DVOs to locate our algorithm's picks in comparison with DVO picks: roughly two-thirds of our initial picks were "vetted out" (and, hence, not treated). Finally, data issues also impacted our post-period to 6-months in which to measure outcomes (and significantly less than this six months for people selected later in the study).

Given the significant limitations to our field experiment, we designed an additional quasi-experiment created by the existence of geographical police precinct boundaries in New York City. Specifically, we plan to take advantage of the geographic nature of our data to compare otherwise similar individuals across command boundaries where, in one command, a potential victim is treated while in a nearby command a similar victim remains untreated for idiosyncratic reasons. This additional analysis would allow us to understand whether individuals chosen by the algorithm are at higher baseline risk than those chosen by officer, one of the intended goals of the RCT. To do so, we will be comparing those individuals who receive home visits because they reside in a particular NYPD command with similar individuals who do not receive home visits because they happen to reside in another NYPD command. This comparison would allow us to create counterfactual outcomes for victims receiving home visits. If these estimates are significantly lower than the realized risk of victims chosen by our model, this difference provides evidence that the algorithm is selecting individuals who are at greater baseline risk than those chosen by officers.

While we have made significant progress in completing both our RCT analysis and our quasi-experiment analysis, the COVID-19 pandemic has triggered an extenuating circumstance that has delayed the completion of these analyses. More specifically, in March 2020 we lost access to our project data, which for security and privacy considerations was required to be stored and analyzed onsite at NYPD headquarters and became inaccessible to us due to the public health emergency.

While our analyses are currently incomplete, our early results convince us that there is value in concluding our study for policymakers, law enforcement agencies and victim service organizations nationwide. Although this grant is ending, we intend to complete the research. Once our study is complete, we still intend to seek to publish the results in a leading peer-reviewed scientific journal, through presentation at a leading computer science conference, and to complement our scientific publications with webinars and other policymaker and media outreach activities.

IV. *Outcomes*

a. Activities/accomplishments

In parallel with the process of developing the prediction model, the research team created Name Match, a probabilistic record linkage toolkit developed using data from the Chicago Police Department, which we believe could be applied to the NYPD data and future phases of the DV prediction model project.. The following activities are associated with Name Match:

- Open sourcing: The code for linking will be open sourced in the second half of 2022
- Publications:
    - Tahamont, S., Jelveh, Z., Chalfin, A., Yan, S., & Hansen, B. (2021). Dude, where's my treatment effect? errors in administrative data linking and the destruction of statistical power in randomized experiments. *Journal of Quantitative Criminology*, *37*(3), 715-749.
    - Tahamont, S., Jelveh, Z., Chalfin, A., Yan, S., & Hansen, B. (2022). No Ground Truth? No Problem: Improving Administrative Data Linking Using Active Learning and a Little Bit of Guile. Under review *PLoS One*

.

b. Results and findings

None to report at this time.

c. Limitations

The primary limitation of our work is that, as a result of the Covid-19 pandemic, we were not able to finalize the analysis prior to the end of the grant period. We intend to complete the work after the grant period.

In considering the ultimate results, and another important limitation is that patterns of domestic violence that exist in New York City may differ from those in other parts of the country potentially limiting generalizability for policymakers and researchers in other jurisdictions.

V. Artifacts

None at this time.