Online Appendix for "Machine Learning Can Predict Shooting Victimization Well Enough to Help Prevent It"

Sara B. Heller, Benjamin Jakubowski, Zubin Jelveh & Max Kapustin

A Methods

This appendix provides additional details regarding our modeling process. First, we describe the raw CPD data, as well as the record linkage algorithm used to identify unique people across records. Next, we discuss the features (predictors) generated from these records. Finally, we discuss how we construct a set of cohorts and use them for model training and evaluation.

A.1 Data

Our model predicts a person's probability of becoming a shooting victim in the 18 months following the date of prediction, using information from 12.7 million CPD records. Available records describe 3,783,724 arrest, 8,911,412 victimization, 23,859 shooting, and 11,598 homicide events that occurred between August 1999 and October 2019 in Chicago.[32] We proceed to describe the relevant attributes of each of type of event record.

A.1.1 Arrest records

CPD arrest records include a unique person identifier (an *Illinois Record (IR) number*), based on a fingerprint scan, that allows us to construct a person's entire CPD arrest history. In addition to this person identifier, CPD arrest records include an incident identifier that can be used to link together the arrestees and victims associated with a single incident. The arrest data also contain police-recorded information about: arresting charges,[33] charge descriptions, and UCR codes; the location and time of the incident and the arrest; demographics of the arrestee; and information about whether the arrest was gang related (and, if so, the arrestee's CPD-identified gang affiliation).

A.1.2 Victim records

While victimization records include an incident identifier, they do not include a unique person identifier. As such, the raw data do not allow us to construct a person's entire victimization history. However, the records provide each victim's identifying information (including name, home address, and date of birth), and therefore support probabilistic

[32] The shooting data start in 2010.

[33] Arrests are associated with one or more arresting charge, and our arrest features consider the full set of charges on the arrest.

record linkage, described below. The victimization records also contain information about the type of victimization incident, including a description and UCR code.

A.1.3 Shooting and homicide victimization records

Similar to victimization records, shooting and homicide victimization records include an incident identifier but no unique person identifier. These records also include demographic information, facilitating record linkage. Shooting and homicide records are used to construct our outcome.

A.2 Record linkage

While CPD arrest records include a unique person identifier, victimization, shooting, and homicide records do not. As such, we use a probabilistic record linkage algorithm to associate unique individuals with all of their records across the CPD data. For details on the algorithm itself, see McNeill and Jelveh (2021). In this section, we describe the basics of the linking procedure.

To link CPD records that refer to same person, we take the post-2010 IR number (the person identifier associated with arrest records) as ground truth, allowing us to identify the set of unique individuals arrested during the study period and to associate these individuals with their arrest records.[34] Since records are already linked within the arrest data, probabilistic record linkage primarily allows us to address two remaining data challenges: associating arrested individuals with their non-arrest records—the various victimization incidents—and identifying the unique individuals represented across the victimization, shooting, and homicide data who did not experience a CPD arrest during the study period.

Our record linkage algorithm produces a collection of records referring to the same person which we call a *cluster*. In assigning records to clusters, the algorithm follows researcher-specified rules based on the context of the data. For our linkage, we specify the following constraints. First, a cluster can have at most one post-2010 IR number. Second, a homicide record cannot link to another record if the homicide record's event date came before the other record's event date. Third, 73.2 percent of victimization records do not have date-of-birth information—an important predictor of true positive links—which can lead to a large number of false positive links. To reduce the chance of these false positives, we introduce a constraint that if at least one record in a record pair is missing date-of-birth information, enforce that the age field (if not missing) in the two records is within 3 years. We also enforce that if at least one record in a potential cluster is missing date-of-birth information, all other records in the cluster not missing date-of-birth information must have similar dates of birth.[35]

Our record linkage procedure identifies 5,426,703 individuals across the three decades of our data. We filter the set of clusters before training and evaluating our predictive models (see details in Appendix A.4.1).

[34] The consistency of IR numbers is somewhat spotty at the beginning of the records but improved considerably over time. As such, we do not treat IR numbers prior to 2010 as ground truth.

[35] We operationalize this by enforcing that these dates of birth be within two character edits of each other.

A.3 Feature generation

Record linkage identifies the set of unique individuals represented in the CPD data, and associates each individual with their CPD arrest, victimization, shooting, and homicide record set. To predict an individual's risk of being shot as of a given prediction date, we aggregate over these associated records to construct *(person, prediction date)*-level features.[36] We construct four broad types of features: demographic, arrest, victimization, and network features. Table A.1 provides a summary of this final feature set, described by type below.

When an individual has no data in either the arrest or the victimization records, we assign a count of 0 to each relevant set of features. For the time-since features, which are not counts, we assign a missing value to the relevant features rather than a 0, and program the LightGBM package to include those instances and count their features as missing. Similarly, when a categorical feature is missing (e.g., police beat or gender), we assign a special category which is treated as missing. If an individual is missing network features due to having no co-arrests or co-victimizations, we assign 0s for those features and include an indicator that the set of those features is missing (i.e., the person is not part of the network map).

Table A.1: Feature counts by type and subtype

Feature Type	Feature Subtype	Count
Demographics	Age	4
Demographics	Race	3
Demographics	Gender	3
Demographics	Police Beat	3
Arrest	Indexed	104
Arrest	Fine-Grained	365
Arrest	Gang	3
Victimization	Indexed	84
Victimization	Fine-Grained	233
Network	1st and 2nd Degree	598
Network	Centrality, degree	10
Total		1,406

A.3.1 Demographic features

We construct 13 demographic features from information on an individual's age, race, gender, and home address.[37] As with most administrative data, police records are often noisy, with different values of theoretically invariant characteristics appearing across multiple records for the same individual. We represent age and race using the modal value across

[36] When generating features for a given *(person, prediction date)*, we restrict to records available prior to the prediction date.

[37] For discussion regarding the inclusion of race in the model, see Appendix B.3.2.

an individual's record set. When exact date of birth is missing, we treat the age feature as missing and construct a missing indicator; this occurs only for 10,766 people who are only in the victimization records (i.e., have never been arrested). However, most of these records include an approximate age, which we use to construct an additional approximate age feature for each individual, as well as a similar missing indicator for approximate age information. [38] We represent gender using three separate features: an individual's (1) most recently recorded gender, (2) modal gender, and (3) the number of distinct genders with which they are associated. We summarize a person's home address and race using these same three types of features.

A.3.2 Arrest features

We construct 472 features summarizing an individual's prior arrest history. These arrest features fall into three broad types: indexed arrest features, fine-grained arrest features, and gang features.

To compute indexed arrest features, we bucket the charges associated with an arrest into several broad, overlapping categories: domestic incidents, drug crime, drug dealing, gun assault or battery, gun battery, gun robbery, property crime, violent crime, Part I violent crime, and all types of crimes. Then, we summarize individual arrest histories within each index using three types of time-aware features:

1. Time since first indexed arrest;

2. Time since most recent indexed arrest;

3. Total number of indexed arrests within the following time windows: the previous 30, 60, 90. 180, 270, 365, or 730 days, and over the individual's entire CPD arrest history (beginning in August 1999).

While these indexed arrest features provide a rich summary of an individual's arrest history, they could still potentially mask heterogeneity in the predictive value of different sorts of incidents collapsed into each index. As such, we augment our representation of prior arrests with 365 fine-grained arrest features that count how many arrests an individual has, within each time window, by unique UCR code and charge.

Finally, in addition to indexed and fine-grained prior arrest features, we compute three measures of an individual's prior CPD-identified gang affiliation. These measures include (i) an indicator of whether the individual has any prior gang-affiliated arrests, (ii) the number of unique gangs with which an individuals has been associated, and (iii) the most recent gang with which an individual is associated.

A.3.3 Victimization features

We construct 317 features summarizing an individual's history of victimization. Paralleling our treatment of prior arrests, we compute both indexed and fine-grained measures of prior victimization, using the same time windows and indices.

[38] We combine true and approximate age information to classify people as over- or under-23 when reporting performance metrics.

A.3.4 Network features

Since CPD arrest, victimization, shooting, and homicide records all share an event identifier, we construct a network using information on events within five years of the prediction date that includes two types of links: (i) links between co-arrestees, and (ii) links between arrestees and victims.[39] After constructing this network, we generate two types of features summarizing an individual's position within it.

First, we compute aggregate statistics describing an individual's network connections (whom we also refer to as neighbors). We compute two types of aggregate statistics. The first counts incidents, while the second counts people. Specifically, the first type of aggregate counts the number of incidents involving an individual's neighbors, by incident type and time window. For example, we count the number of property crime incidents that occurred in the last 365 days and resulted in the arrest of a neighbor. The second type of aggregate counts the number of neighbors involved in incidents, again by incident type and time window. For example, we count the number of neighbors arrested for property crime incidents within the last 365 days. We compute these two types of aggregates separately for an individual's first- and second-degree network connections.

Second, we compute features describing the underlying network structure, including an individual's degree and eigenvector centrality, as well as the maximum degree and eigenvector centrality of their first- and second-degree neighbors.

A.4 Model training

To maximize flexibility, especially in the top tail of the risk distribution, we train and test a gradient-boosted decision tree model (Friedman, 2002). Because we include network features in the model, we cannot use traditional sub-sampling to generate a hold-out test set. Even if individual i were part of a randomly sub-sampled test set, information about i's risk could still be used in model-building to the extent he has peers in the training data and appears in their features. To avoid this kind of overfitting, we divide the data into calendar time cohorts as follows.

A.4.1 Defining cohorts

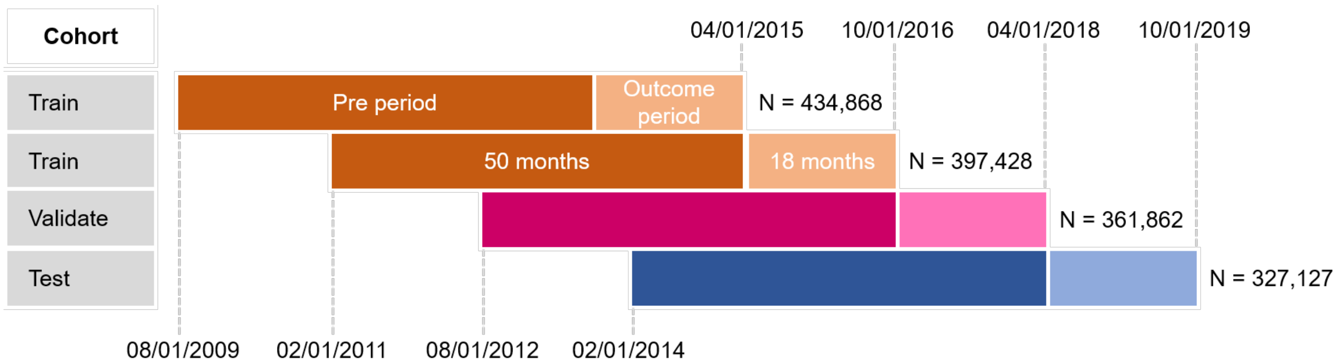
To define cohorts, we first establish four non-overlapping 18-month outcome periods. Then, we identify cohorts of individuals who have had either an arrest or two victimizations during the 50 months preceding each 18-month outcome period (see Figure A.1). This cohort definition excludes two sets of people: those with no CPD records from the past 50 months, and those who only had a single victimization in the past 50 months. We exclude these individuals for two reasons. First, they have much lower baseline risk: 0.01 percent of these individuals were shot during the follow-up period, compared to 0.7 percent of people in our cohort. Second, and as we describe above, we are more conservative in allowing links for victimization records without date-of-birth information. This leads to large number of singleton victimization records which remain unlinked. Therefore, dropping single victimizations implicitly reduces the influence of record-linkage error

[39] Note this corresponds to the bipartite projection of the bipartite $person \leftrightarrow incident$ graph.

caused by poor data quality. As such, our sample selection criteria reduce data integrity issues, while still capturing most of the identifiable population with elevated risk.

We use the first two cohorts to train the model. We split the third cohort into a 50 percent validation set for hyperparameter tuning, a 25 percent set for calibrating the predictions from the model, and a 25 percent set to optimize the number of trees in the gradient boosting model via "early stopping" (Raskutti et al., 2011).[40] The final cohort is our test set, where we predict shooting risk ($\hat{p}$) for the out-of-sample 18-month outcome period starting on April 1, 2018.

Figure A.1: Model cohort structure



A.4.2 Hyperparameter tuning

We optimize the performance of our gradient-boosted decision tree model using random search over the following hyperparameters: number of leaves, minimum number of data in each leaf, learning rate, and the fraction of data instances and features to use in building each tree. Our random search procedure is as follows:

1. We randomly sample $N = 100$ hyperparameter configurations from this search space.

2. For each hyperparameter configuration, we fit a gradient-boosted decision tree model over the two training cohorts, using early stopping (based on minimizing log loss on a partition of the validation cohort) to optimize the number of rounds of boosting (i.e., the number of decision trees in the ensemble).

3. From this set of $N = 100$ random hyperparameter configurations, we select the configuration that maximizes precision evaluated at the rank that equals the number of shooting victims in the validation set.

4. Finally, we refit the model, using the selected hyperparameters, over the combined training and validation cohorts.

[40] In this paper, we only report the results for the raw predictions, and not the predictions that are rescaled by the calibration model.

A.5 Model evaluation

We evaluate the performance of our model on the test set (prediction sample). While our primary evaluation metrics are described in the main paper, this section provides additional detail on (i) construction and interpretation of the $\hat{p}$-weighted prediction sample (Figure 3) and (ii) construction of bootstrap confidence intervals for precision at k (Figure 4 and Appendix Figure B.4).

A.5.1 $\hat{p}$-weighted prediction sample

Figure 3 includes a "Predicted victims" series that shows the demographic composition of a weighted sample, where individuals in the prediction sample are weighted based on their predicted shooting risk $\hat{p}_i$. Specifically, for a given demographic subgroup G, this series shows

$$\%\text{ in demographic group } G = \frac{\sum_{i \in G} \hat{p}_i}{\sum_i \hat{p}_i}$$

where $\hat{p}_i$ is predicted risk for the i^{th} individual. If the model generated perfect predictions, then the demographic composition of predicted victims would be the same as the demographic composition of actual victims in the prediction sample. As such, differences between the second and third horizontal bars in Figure 3 indicate misprediction.

A.5.2 Bootstrap confidence intervals

Tables B.1 and B.5, as well as Figures 2, 4, and Appendix Figure B.4, include 95 percent bootstrap confidence intervals for several statistics at different k. These are constructed from 1,000 bootstrap samples, where each bootstrap sample is generated by:

1. Bootstrap resampling the prediction sample (i.e., drawing $N_{prediction} = 327,127$ instances from the test set, with replacement).

2. Within each bootstrap sample, computing Precision_k and Recall_k at different k (e.g., $k = 1, 2, \ldots, 5000$).

The 95 percent confidence intervals report the 2.5^{th} and 97.5^{th} percentiles from this bootstrap distribution.

While this bootstrap procedure characterizes prediction set sample variance, it does not account for other sources of variation in our procedure (e.g., training set sample variance, or explicit randomness in the gradient boosting algorithm).

B Additional Results

B.1 Prevalence of other outcomes among those with high predicted risk of shooting victimization

The main text reports predictive performance for the primary outcome of interest, shooting victimization, when ranking people by their predicted risk of that outcome. Because the risk of being shot is likely correlated with the risk of other socially costly outcomes, efforts to reduce the risk of shooting victimization among this group may reduce the risk of these other outcomes as well. We do not focus on quantifying the benefits of reducing the risk of these other outcomes, since they are less reliable measures of the underlying behavior of interest (i.e., the relationship between arrest for violent crime and true violent offending is likely to be noisier and to differ by racial group, relative to the relationship between shooting victimization in the police data and true shooting victimization).

Nonetheless, because efforts to prevent shooting victimization among this group may produce other large benefits, this section reports on the prevalence of other measures of violence among those predicted to be at high risk of shootings. Note that we are not training a model to predict these other outcomes, since that would likely confound police behavior or willingness to report violence to the police with true individual risk. Rather, we are reporting on the prevalence of different violence measures among groups defined by their ranking in the shooting victimization predictions.

Appendix Table B.1 below reports our standard measures of model performance, precision and recall, for the full shooting victimization model evaluated on four different outcomes: shooting victimization, shooting arrest, violent crime victimization, and violent crime arrest.

Table B.1: Predictive performance of shooting victimization predictions for other outcomes

k	Precision	Recall	Total Recall
Shooting Victim			
500	0.130 (0.100, 0.160)	0.029 (0.022, 0.036)	0.019 (0.015, 0.024)
3,381	0.087 (0.078, 0.097)	0.130 (0.117, 0.146)	0.087 (0.078, 0.097)
327,127	0.007 (0.007, 0.007)	1.000 (0.958, 1.041)	0.666 (0.638, 0.694)
Shooting Arrest			
500	0.034 (0.018, 0.050)	0.047 (0.025, 0.070)	0.037 (0.019, 0.054)
3,381	0.020 (0.016, 0.025)	0.192 (0.148, 0.234)	0.148 (0.114, 0.181)
327,127	0.001 (0.001, 0.001)	1.000 (0.900, 1.097)	0.772 (0.695, 0.847)
Violent Crime Victim			
500	0.214 (0.178, 0.250)	0.007 (0.005, 0.008)	0.003 (0.002, 0.003)
3,381	0.170 (0.158, 0.182)	0.035 (0.032, 0.038)	0.014 (0.013, 0.015)
327,127	0.050 (0.049, 0.051)	1.000 (0.985, 1.015)	0.397 (0.391, 0.402)
Violent Crime Arrest			
500	0.178 (0.148, 0.216)	0.018 (0.015, 0.022)	0.012 (0.010, 0.014)
3,381	0.151 (0.138, 0.162)	0.103 (0.095, 0.111)	0.068 (0.062, 0.073)
327,127	0.015 (0.015, 0.016)	1.000 (0.974, 1.028)	0.658 (0.641, 0.676)

Note: Performance and recall from the full model trained to predict shooting victimization during the 18-month outcome period starting April 1, 2018. Model performance is evaluated on the four outcomes shown, for the k people with the highest predicted risk of shooting victimization. Violent crime arrest refers to the Part I violent index offenses: aggravated assault, aggravated battery, forcible rape, murder, and robbery. Prediction sample size is 327,127.

The people whom the model predicts to be at higher risk of shooting victimization are indeed at higher risk for these other adverse outcomes during the 18-month outcome period as well. For example, among the 500 people at highest predicted risk of shooting victimization, 3.4 percent are arrested on suspicion of carrying out a shooting (34 times the base rate in the whole test set of 0.1 percent); 21.4 percent are reported as victims of a violent offense (4.3 times the base rate); and 17.8 percent are arrested on suspicion of carrying out a violent offense (11.9 times the base rate).

B.2 Victim counts and performance by demographic group

Figures 2 and 3 in the main text show the proportion of shooting victims that fall into different demographic groups. This section adds some additional information to the summaries in the main text. To be transparent about the underlying size of each group, Appendix Table B.2 below reports the counts across demographic categories of four groups: all shooting victims, shooting victims in the prediction sample, predicted victims (see discussion above in Appendix A.5.1), and the $k = 3,381$ people with the highest predicted risk.

Table B.2: Demographic composition of actual and predicted shooting victims

Race	Gender	Age	Actual victims (N=3381)	Actual victims (in sample) (N=2253)	Predicted victims (rounded)	Top 3,381
Black	Male	<23	885	672	579	1976
		23+	1456	1072	966	1031
	Female	<23	159	65	39	1
		23+	214	123	98	4
Hispanic	Male	<23	210	112	114	236
		23+	276	152	188	121
	Female	<23	22	10	9	0
		23+	47	15	19	0
White	Male	<23	10	6	8	4
		23+	46	19	29	8
	Female		13	6	11	0
Other/Missing			43	1	9	0

Note: Counts for White females of all ages reported due to small cell sizes.

Figure 1 in the main text shows that the predictions are well-calibrated overall and by racial group, with some under-prediction in the high-risk tail of the distribution for White and Hispanic individuals but over-prediction in the last bin. Appendix Table B.3 sheds additional light on calibration by contrasting the base shooting victimization rate within the prediction sample and the average prediction, both by race/ethnicity (as in Figure 1) and further broken down by age and gender (as in Figure 3).

Consistent with the calibration plots in the main text (Figure 1), average predictions are generally quite similar to observed rates of shooting victimization, even within race/ethnicity-age-gender groups. The model slightly under-predicts risk for Black men and women (by anywhere from 0.001 for older Black men and women to 0.004 for younger Black men), while it slightly over-predicts risk for older Hispanic men (by 0.001) and older White women.

Table B.3: Base rate and average predicted risk by race, gender, and age for prediction sample

Race	Gender	Age	N	Base Rate by Group	Mean predicted risk
Black	All	All	199192	0.01	0.008
	Female	All	77371	0.002	0.002
		<23	12141	0.005	0.003
		23+	65230	0.002	0.001
	Male	All	121782	0.014	0.013
		<23	18939	0.035	0.031
		23+	102842	0.01	0.009
Hispanic	All	All	68613	0.004	0.005
	Female	All	20333	0.001	0.001
		<23	3442	0.003	0.003
		23+	16890	0.001	0.001
	Male	All	48260	0.005	0.006
		<23	7639	0.015	0.015
		23+	40619	0.004	0.005
White	All	All	49710	0.001	0.001
	Female	All	18443	<0.001	0.001
		<23	1258	0.001	0.001
		23+	17183	<0.001	0.001
	Male	All	31235	0.001	0.001
		<23	2281	0.003	0.003
		23+	28943	0.001	0.001
Other Race/Gender		All	7287	<0.001	0.001
Missing Race/Gender/Age		All	2434	<0.001	0.001

Note: Table shows the base rate, or the proportion of each group that becomes a shooting victim during the outcome period, along with the average predicted risk within each group. Note that the "All" age rows include individuals of that race/ethnicity and gender who are missing age information; as a result, the number of observations in the under- and over-23 rows do not exactly total to N for the "All" row. The final row groups everyone with missing race/ethnicity, gender, and/or age information together.

Of course, average predictions being similar to base rates at a group level does not mean each individual's prediction is accurate. To assess accuracy at the individual level, one must establish a decision rule that translates predicted risk levels into classifications of "positive" (predicted to be shot) and "negative" (predicted not to be shot) for each person. There are many different classification rules one could use. Given the uneven demographic distribution of individuals across the risk distribution, different decision rules could have different implications for who is correctly and incorrectly classified.

Since a natural kind of decision rule is a threshold rule, where policymakers would consider everyone above some global risk threshold as a positive prediction and everyone below as a negative prediction, we show the implications of one such threshold (the same that is shown in Figure 3): serving the 3,381 individuals with the highest predicted risk (motivated by the fact that there are 3,381 actual victims in the outcome period). Appendix Table B.4 shows precision and average predicted risk within race/ethnicity and age groups for the subset of men among the top 3,381 highest predictions. We omit women and the age breakdown for White men in this table because there are so few of these individuals in this top-ranked group.

Table B.4: Precision and average predicted risk by race and age for men among the top 3,381

Race	Gender	Age	N	Precision	Mean predicted risk
Black	Male	All	3007	0.092	0.102
		<23	1976	0.104	0.106
		23+	1031	0.069	0.096
Hispanic	Male	All	357	0.048	0.104
		<23	236	0.055	0.103
		23+	121	0.033	0.105
White	Male	All	12	0.0	0.198

Note: Table reports statistics for White males of all ages together and omits 6 individuals belonging to other demographic groups due to small cell sizes.

Comparing the two columns gives a sense for subgroup calibration for this subsample, and precision shows the proportion of true positives (such that $1 - Precision$ is the false discovery rate). Again we emphasize that this not reflective of performance across the whole sample, but rather provides additional information on the fairness implications of a "top 3,381" decision rule.[41]

Comparing the mean predicted risk with the realized risk (precision) in Appendix Table B.4 shows several key patterns. First, consistent with the subgroup calibration panels in Figure 1, predicted risk among this top tail is quite close to the realized risk for Black men, but overstates the realized risk for Hispanic men on average. The age breakdown further shows that the predictions are best calibrated for younger minority men, who tend to have fewer but more recent police contacts. The older minority men chosen with this decision rule tend to have elevated predictions relative to their realized risk.

In terms of classification among the top 3,381, the model has the highest true positive rate (and thus lowest false discovery rate) for Black men, of whom 9.2 percent are correctly classified, i.e., become shooting victims in the outcome period. In contrast, among Hispanic men—a much smaller group of 357 compared to 3,007 Black men—only 4.8 percent are correctly classified. This is consistent with argument in the main text that the over-representation of Black men in the top tail of the risk distribution is not because estimates of their risk are distorted (inflated), but rather because the model does a better job at identifying Black men who face genuinely higher risk of victimization. These true positive rates are extremely high from a substantive standpoint, identifying 276 Black men and 17

[41] The reported numbers for White men have huge implicit confidence intervals since there are only 15 of them; while it is notable that none of these 15 is shot during the outcome period, we hesitate to over-interpret a pattern from so few individuals.

Hispanic men for whom preventive services might have kept them from serious injury or death. Nonetheless, the fact that 91 percent of Black men and 95 percent of Hispanic men above this threshold are not shot during the outcome period again emphasizes how costly it would be target any intervention that reduced people's civil liberties based on these predictions.

B.3 Further detail on what matters for prediction

B.3.1 Performance by groups of features

The main text presents predictive performance leaving out 3 sets of features: own arrests, peer information (networks), and both. We perform a similar exercise, dropping sets of features and retraining a new model, for additional combinations of features. Appendix Figure B.1 reports precision for the full model and different models that each exclude certain feature sets. To ensure the lines are not all on top of each other, we limit the scale to the top 5,000 ranked individuals in each model. Past 5,000, most of the differences in performance tend to be quite small. We do not show confidence intervals to make the figure more readable, but note that there is a fair amount of noise from sampling variation at any given k. Appendix Table B.5 quantifies the precision differences at $k = 500$ and $k = 3,381$, as well as reporting recall and total recall.

Figure B.1: Precision across models with different feature sets

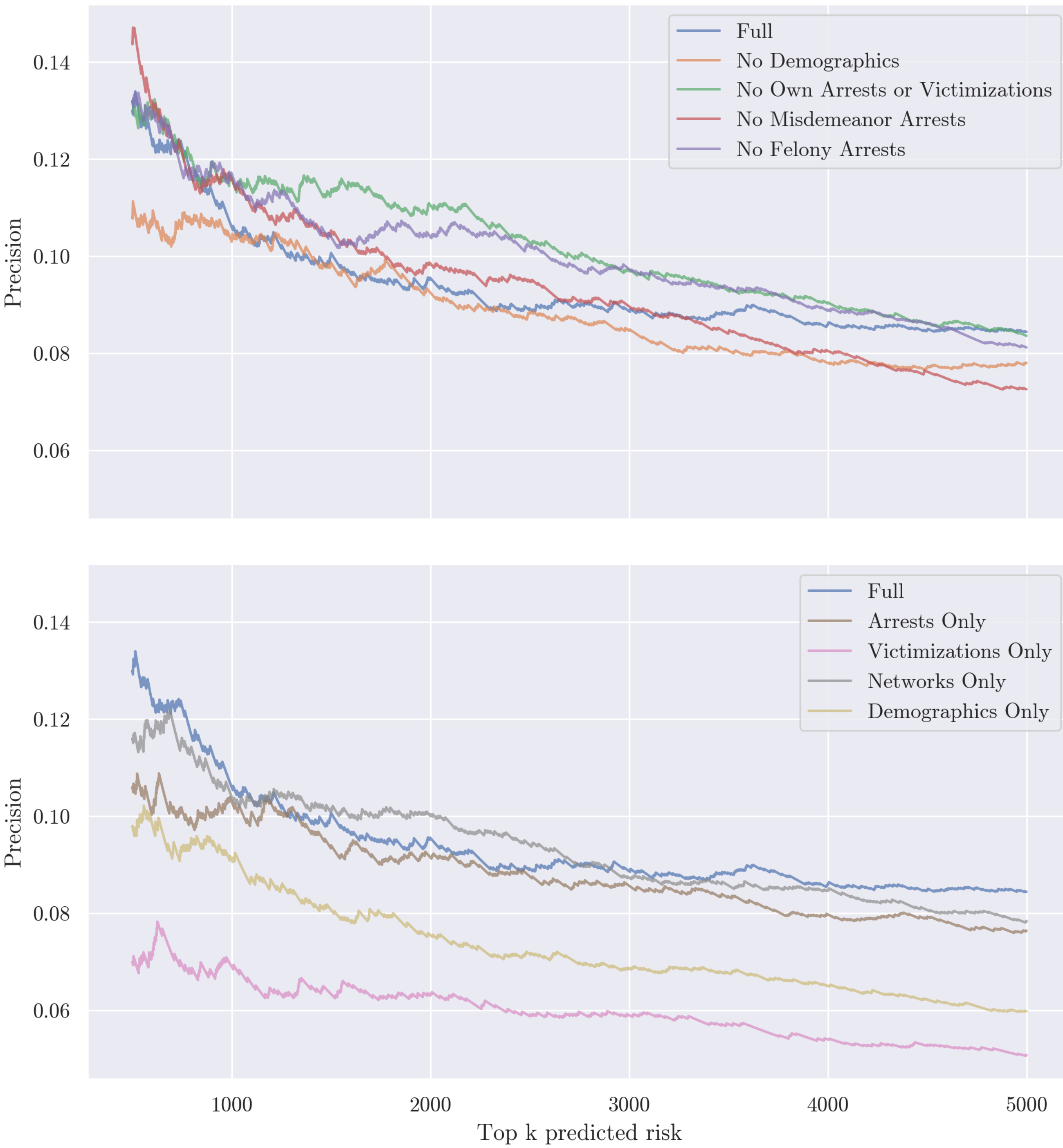


Table B.5: Predictive performance by feature set

	Top 500			Top 3,381		
Feature Set	Precision	Recall	Total Recall	Precision	Recall	Total Recall
Full	0.130 (0.102, 0.154)	0.029 (0.023, 0.034)	0.019 (0.015, 0.023)	0.087 (0.079, 0.098)	0.130 (0.118, 0.147)	0.087 (0.079, 0.098)
No Network Information	0.132 (0.103, 0.162)	0.029 (0.023, 0.036)	0.020 (0.015, 0.024)	0.093 (0.083, 0.103)	0.139 (0.124, 0.154)	0.093 (0.083, 0.103)
No Own Arrests	0.126 (0.099, 0.155)	0.028 (0.022, 0.034)	0.019 (0.015, 0.023)	0.091 (0.082, 0.102)	0.137 (0.122, 0.153)	0.091 (0.082, 0.102)
No Own Arrests, No Network Information	0.098 (0.073, 0.124)	0.022 (0.016, 0.028)	0.014 (0.011, 0.018)	0.075 (0.066, 0.084)	0.112 (0.099, 0.126)	0.075 (0.066, 0.084)
No Race	0.114 (0.090, 0.141)	0.025 (0.020, 0.031)	0.017 (0.013, 0.021)	0.093 (0.085, 0.104)	0.140 (0.127, 0.157)	0.093 (0.085, 0.104)
No Demographics	0.108 (0.086, 0.136)	0.024 (0.019, 0.030)	0.016 (0.013, 0.020)	0.081 (0.073, 0.089)	0.122 (0.110, 0.134)	0.081 (0.073, 0.089)
No Own Arrests or Victimizations	0.132 (0.109, 0.173)	0.029 (0.024, 0.038)	0.020 (0.016, 0.026)	0.095 (0.086, 0.106)	0.142 (0.129, 0.158)	0.095 (0.086, 0.106)
Arrests Only	0.106 (0.079, 0.137)	0.024 (0.017, 0.030)	0.016 (0.012, 0.020)	0.085 (0.076, 0.094)	0.127 (0.114, 0.140)	0.085 (0.076, 0.094)
Victimizations Only	0.070 (0.053, 0.089)	0.016 (0.012, 0.020)	0.010 (0.008, 0.013)	0.057 (0.051, 0.068)	0.086 (0.077, 0.102)	0.057 (0.051, 0.068)
Demographics Only	0.098 (0.075, 0.126)	0.022 (0.017, 0.028)	0.014 (0.011, 0.019)	0.069 (0.060, 0.077)	0.103 (0.090, 0.116)	0.069 (0.060, 0.077)
Networks Only	0.116 (0.097, 0.148)	0.026 (0.021, 0.033)	0.017 (0.014, 0.022)	0.087 (0.078, 0.098)	0.130 (0.118, 0.147)	0.087 (0.078, 0.098)
Arrests + Networks Only	0.124 (0.095, 0.155)	0.028 (0.021, 0.034)	0.018 (0.014, 0.023)	0.091 (0.083, 0.101)	0.137 (0.124, 0.152)	0.091 (0.083, 0.101)
No Own Victimizations	0.116 (0.088, 0.144)	0.026 (0.020, 0.032)	0.017 (0.013, 0.021)	0.093 (0.085, 0.105)	0.140 (0.128, 0.158)	0.093 (0.085, 0.105)
No Misdemeanor Arrests	0.144 (0.117, 0.180)	0.046 (0.037, 0.057)	0.021 (0.017, 0.027)	0.086 (0.076, 0.095)	0.184 (0.164, 0.203)	0.086 (0.076, 0.095)
No Felony Arrests	0.132 (0.109, 0.168)	0.035 (0.029, 0.045)	0.020 (0.016, 0.025)	0.094 (0.086, 0.103)	0.171 (0.156, 0.187)	0.094 (0.086, 0.103)

Note: Models differ based on the feature sets available to them during training (see text below). Model performance is evaluated on shooting victimization during the outcome period for the $k = 500$ and $k = 3,381$ people with the highest predicted risk of shooting victimization.

The definitions of the models that leave out particular feature sets are as follows:

Top panel

1. Full: The main model reported in the text with all available features
2. No Own Victimizations: Excludes all victimization records for the focal person (but includes them for first- and second-degree peers)
3. No Demographics: Excludes race, gender, age, and location information
4. No Own Arrests or Victimizations: Excludes both arrest and victimization features for the focal person (but includes them for first- and second-degree peers)
5. No Misdemeanor Arrests: Excludes all misdemeanor arrests when constructing features, altering counts for focal person as well as first- and second-degree peers
6. No Felony Arrests: As above but excludes felony arrests and includes misdemeanor arrests

Bottom panel

1. Full: Same as above
2. Arrests Only: Uses only information on arrests, excluding victimization and demographic information
3. Victimizations Only: Uses only information on victimizations, excluding arrests and demographic information
4. Networks Only: Uses only information on first- and second-degree peers, excluding all information on the focal person
5. Demographics Only: Uses only information on demographics, excluding arrests and victimization information

Across models that include some version of arrest information (top panel), it appears that different features are correlated enough that the information lost when one set of features is excluded can be replaced by the information in the remaining sets of features. Although some of the prediction gaps between these models could be substantively important in practice—on the order of 2 percentage points, with felony arrests containing the most information—we cannot distinguish the differences from statistical noise.[42] Even the imprecise differences are concentrated at the very top of the risk distribution; after about $k = 3,000$, precisions across most models are basically on top of each other.

As the bottom panel shows, the biggest loss of information comes from using only victimization records when building the model. Using just demographics or a combination of demographics and victimization records does slightly better than victimizations alone, but not as well as other models. This pattern echoes the finding in the main text that the information contained in arrests records (both of one's self and one's peers) is particularly valuable in estimating the risk of being a shooting victim.

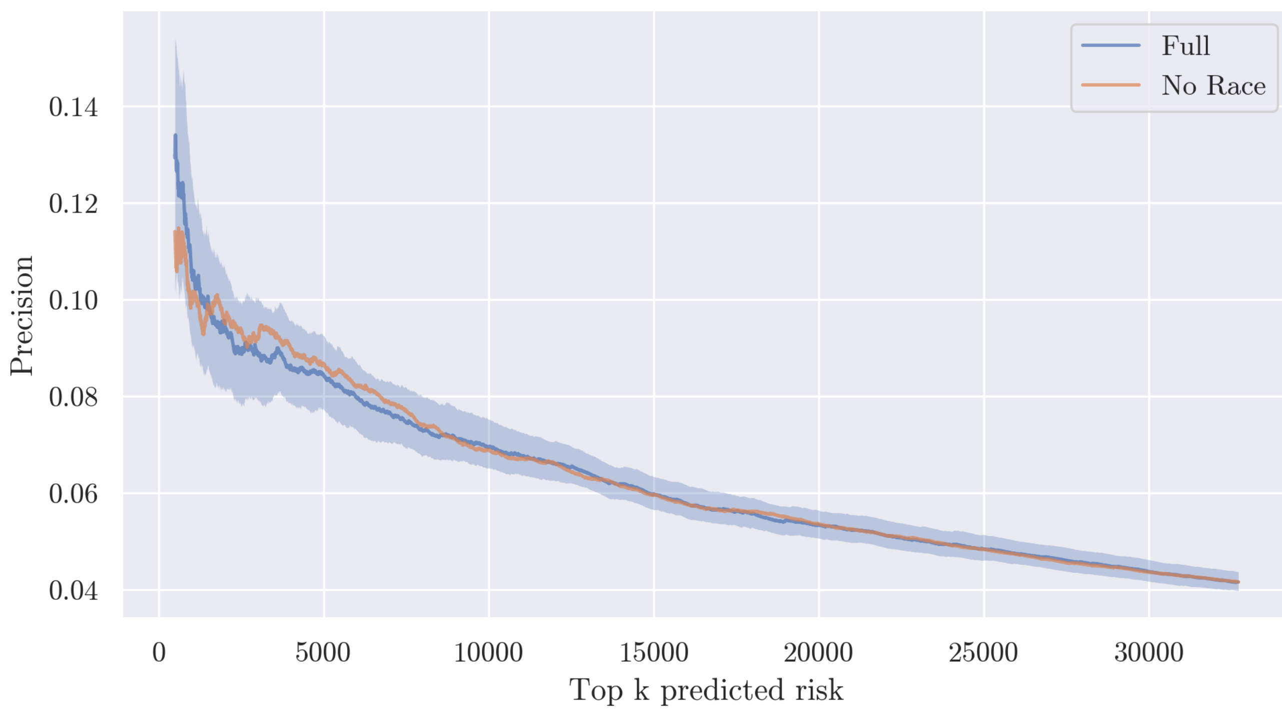
[42] Noise helps to explain why there is a small gap above $k = 1,500$ where the "no own arrest" model appears to outperform the full model.

B.3.2 Prediction without race

Our main results come from a model that includes race in the model-building process. Many legal scholars believe that including race as an algorithmic input is likely unconstitutional, though the debate around this question is not completely settled (e.g., Yang and Dobbie, 2020).

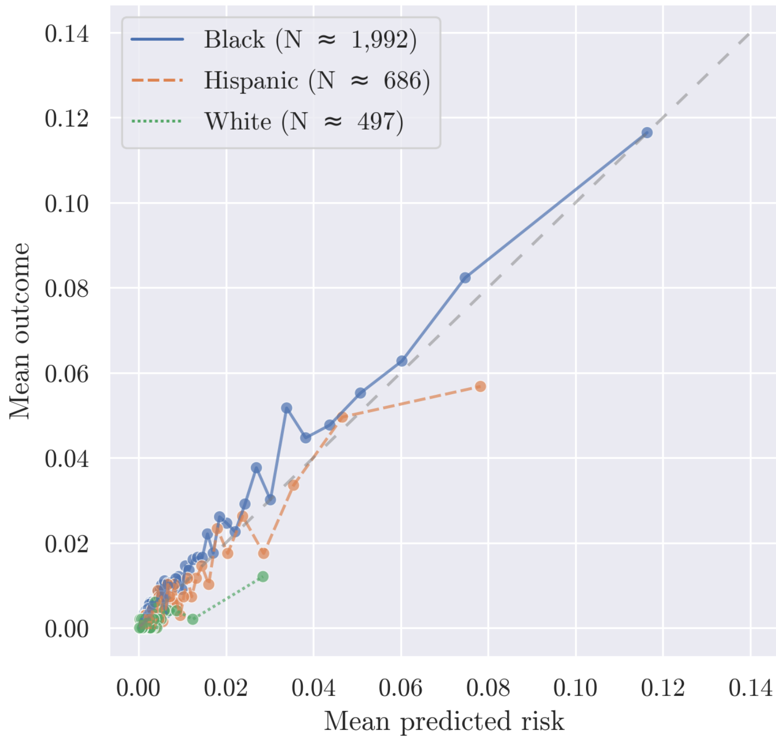
Importantly, as shown in Appendix Figure B.2, the inclusion of race has a relatively trivial effect on predictive performance.[43]

Figure B.2: Precision across models with and without race indicators



[43] We show performance for the top 10 percent of the sample, since showing the full sample makes the scale too small to see the very small differences at the top.

Figure B.3: Calibration for model built without race indicators



There is perhaps some small loss of precision at the extreme top of the risk distribution, but it is not statistically distinguishable from the full model that uses race. This is not unusual in settings where many other features in the prediction are correlated with race (Starr, 2014). Appendix Figure B.3 also shows little change in calibration within race/ethnicity groups relative to the full model shown in Figure 1. So although we show the main results from a model including race, the arguments contained in the paper are equally applicable for settings that require the model to exclude race.[44]

B.3.3 Performance by number of features & model complexity

A different way to ask what information matters is not to focus on sets of features grouped by theme, but on the number of features available and the complexity of the algorithm used to predict with them. Black box models may not be appropriate in all high-stakes settings (Rudin, 2019). A simpler model with only a few features may aid in interpretability, trust, and uptake (Ustun and Rudin, 2019). Multiple researchers have identified settings where complex models with more features provide minimal performance improvements over simple models with fewer features (e.g., Dressel and Farid, 2018; Jung et al., 2017; Angelino et al., 2018; Stevenson and Slobogin, 2018; Stevenson and Mayson, 2021). Thus, for use in these contexts, it is important to understand how much of the predictive accuracy of the

[44] When a somewhat different version of this prediction model was used for social service referrals in practice, we excluded race; see, e.g., https://osf.io/ap8fj/.

full model can be captured by a drastically smaller set of features and simpler modeling techniques.

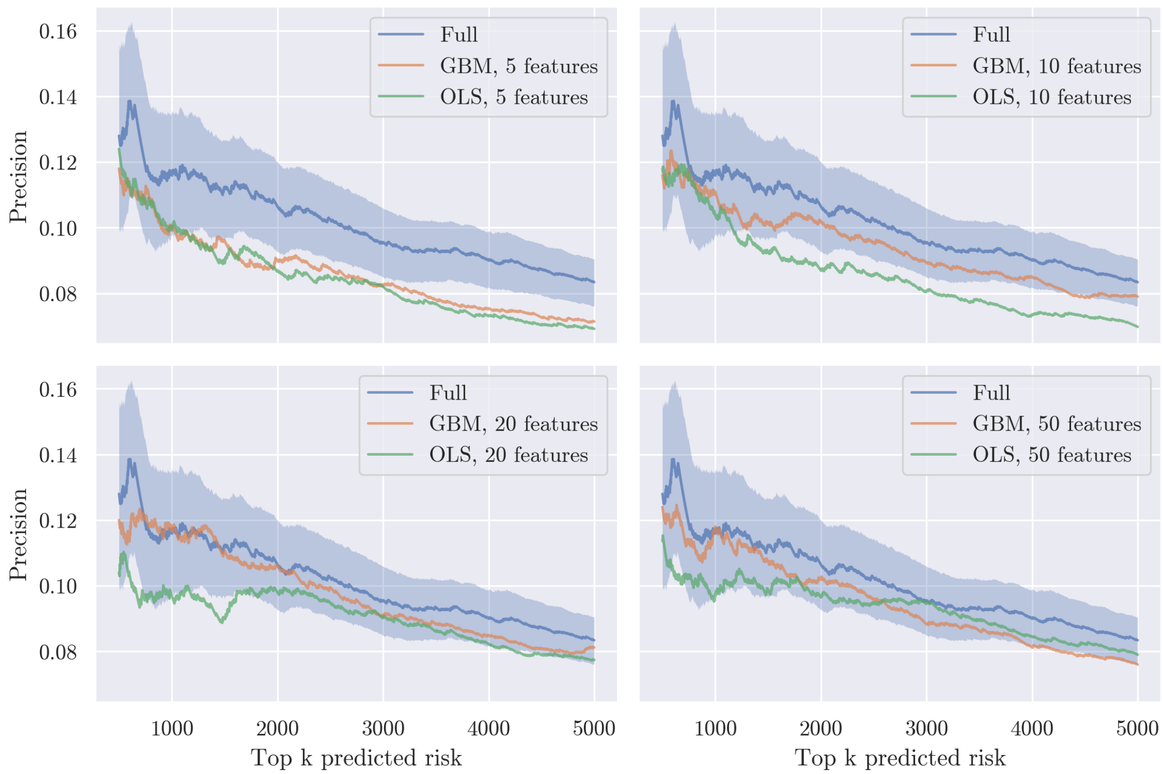
We explore these questions in our setting by first creating a rank-ordered set of 50 features from the full set of all 1,406 features, where the first feature has the highest correlation with shooting victimization, the second has the highest correlation after residualizing out the first feature, and so on. Then we build models using both gradient-boosted decision trees (GBM) and ordinary least squares (OLS) using only the $n \in \{5, 10, 20, 50\}$ highest-ranked features, comparing their performance to that of the full model built using GBM with 1,406 features.

To generate this smaller set of 50 features, we use a simple stepwise residualization procedure. First, we select the single feature that is most highly correlated with shooting victimization in the first two cohorts. Then we remove the correlation between all other features and the selected feature. To do so, we replace the value of each unselected feature with the residual from a linear regression of each unselected feature onto the feature with the highest correlation. Finally, we repeat the process, searching each time for the feature with the highest remaining correlation with the outcome after removing the correlation with already-selected features. Given a particular collection of features to start with, this approach produces a rank ordering of the features in that collection with the highest linearly independent relationship with the outcome.

Appendix Table B.6 reports the set of 50 features chosen by this process. The first column shows the set to which each feature belongs; the second column provides a description of the feature, where text in parentheses indicate a subtype of the feature; the third column shows, where appropriate, the time window over which the feature was measured, where "Total" indicates features that look back to the beginning of the data (August 1999); the fourth column shows the correlation between the residualized version of the feature and the outcome; and the fifth column shows the correlation between the unresidualized version of the feature and the outcome.

Appendix Figure B.4 reports the same precision plot as Figure 4, with separate panels for different numbers of the top n features reported in Appendix Table B.6. Each panel shows the precision for the full model, and for models using GBM and OLS with only the indicated top n features. As the upper left panel of Appendix Figure B.4 shows, though the differences are noisy, models using only the top $n = 5$ features attain somewhat lower precision: 10 percent of the $k = 1,000$ people with highest predicted risk identified by these very parsimonious models are actually shot in the outcome period, compared to almost 12 percent of those identified by the full model. However, as the number of features available to them rises, the performance of both the GBM and OLS models improve, substantially closing the gap with the full model. This gap closes somewhat faster for the GBM model than for the OLS model. Although the differences in precision are not always statistically significant when accounting for sampling variation, the pattern of results suggests that the additional flexibility of GBM could materially improve performance in situations where the number of available features is more limited. When enough information is available, however, less computationally-intensive prediction models perform as well as more flexible algorithms.

Figure B.4: Precision across models with different model types and number of features



Note: Figure shows precision, or the share actually victimized during the outcome period, of the $k \leq 5,000$ people with the highest predicted risk of shooting victimization, for the full model, a gradient-boosted model with a limited set of features, and an ordinary least squares model with the same limited set of features (Appendix Table B.6). Due to noise in precision at low values of k, we start the graphs at $k = 500$. 95 percent bootstrap confidence interval for full model shown (see Appendix A.5.2 for details).

In addition to Appendix Table B.6, we repeat the stepwise residualization procedure described above for the other feature sets shown in Figure 4. Appendix Tables B.7, B.8, and B.9 respectively show the list of the top 50 features identified by this process for the following three feature sets: no network features, no own-arrest features, and the combination of no network and no own-arrest features.

Appendix Table B.10 reports the main performance metrics for these different models: the full model and each leave-out-a-feature-set model. To generate these metrics, we reran our modeling process but only gave the algorithm access to the 50 most correlated variables as per our stepwise residualization procedure. As the feature lists show, the top 50 predictors change quite a bit as different feature sets are removed. But both precision and recall at $k = 500$ and $k = 3,381$ are quite similar across all models that use 50 features. This emphasizes the point in the main text that standard "importance" measures within a single model do not capture which variables are uniquely important to prediction; other correlated variables can often capture similar information when the "important" variables are removed. To get a clearer understanding of which kinds of features truly matter, in the sense that their removal would harm predictive performance, we must compare predictive performance in models trained without particular variables, as in the main text.

Appendix Table B.10 and Appendix Figure B.4 highlight that it is generally possible to achieve comparable performance to the full model in the tail. The biggest loss in performance is from removing all arrest information on both focal individuals and their neighbors. Of course, it is typically impossible to know *a priori* which small set of features will achieve performance as close as possible to a model with access to the full set of features. The process of solving this constrained optimization problem is itself a machine learning challenge (Rudin, 2019). In practice, settings that require smaller numbers of features could engage in this process.[45]

[45] See Luminosity and York (2020) for a real-world example of developing a risk assessment for pretrial arraignment decisions in New York City.

Table B.6: Top 50 features from the stepwise residualization procedure when given access to all feature sets

Feature Set	Description	Time Window	Correlations Residualized	Correlations Original
networks	# of 1st degree neighbors arrested (any)	365 days	0.120	0.120
arrests	Ever gang-affiliated		0.062	0.095
victims	# of victimizations (shootings)	Total	0.037	0.064
arrests	# of own-arrests (any)	730 days	0.031	0.109
demographics	Age (modal)		0.029	-0.052
arrests	# of own-arrests (gambling)	Total	0.024	0.068
arrests	# of own-arrests (violent)	Total	0.021	0.074
networks	# of 2nd degree neighbors arrested (drug)	30 days	0.021	0.025
arrests	# of own-arrests (reckless conduct)	730 days	0.020	0.067
arrests	# of own-arrests (criminal trespass)	270 days	0.016	0.036
networks	# of 1st degree neighbors arrested (drug)	730 days	0.016	0.102
demographics	Sex (most recent)		0.015	-0.017
demographics	Missing date of birth		0.014	-0.014
arrests	# of own-arrests (solicitation)	730 days	0.014	0.055
arrests	# of own-arrests (drug)	Total	0.014	0.049
victims	# of victimizations (shootings)	270 days	0.013	0.040
networks	# of 2nd degree neighbors victimized (gun assault or battery)	60 days	0.012	0.013
arrests	# of own-arrests (solicitation)	Total	0.012	0.060
networks	# of arrests (any) by 1st degree neighbors	Total	0.012	0.097
arrests	# of own-arrests (firearm possession)	Total	0.012	0.035
demographics	# of unique police beats		0.012	0.027
arrests	# of own-arrests (reckless conduct)	Total	0.012	0.072
networks	# of victimizations (gun battery) by 2nd degree neighbors	60 days	0.012	0.006
networks	# of arrests (drug) by 2nd degree neighbors	730 days	0.012	0.057
networks	# of arrests (drug) by 2nd degree neighbors	30 days	0.011	0.032
arrests	# of own-arrests (robbery)	Total	0.011	0.056
networks	# of victimizations (gun battery) by 1st degree neighbors	Total	0.011	0.071
arrests	# of own-arrests (gun assault or battery)	Total	0.011	0.034
networks	# of 1st degree neighbors arrested (property)	Total	0.010	0.105
networks	# of 1st degree neighbors arrested (drug)	Total	0.010	0.114
networks	# of 2nd degree neighbors victimized (gun battery)	60 days	0.010	0.005
arrests	# of own-arrests (property)	730 days	0.010	0.036
arrests	# of own-arrests (obstructing identification)	Total	0.010	0.033
arrests	# of own-arrests (gang loitering)	Total	0.009	0.043
arrests	# of own-arrests (replica firearm)	Total	0.009	0.030
networks	# of arrests (gun battery) by 2nd degree neighbors	730 days	0.009	0.006
arrests	# of own-arrests (burglary)	Total	0.009	0.036
arrests	# of own-arrests (dealing marijuana public school)	Total	0.008	0.035
networks	# of arrests (gun robberies) by 1st degree neighbors	730 days	0.008	0.018
victims	# of victimizations (any victimization)	Total	0.008	-0.017
arrests	# of own-arrests (drug)	730 days	0.008	0.040
arrests	# of own-arrests (gang name indicator)		0.008	-0.014
networks	# of 1st degree neighbors arrested (gun robberies)	Total	0.008	0.044
arrests	# of own-arrests (aggravated assault school employee)	Total	0.008	0.027
arrests	# of own-arrests (weapons possession)	Total	0.008	0.031
arrests	# of own-arrests (gun robberies)	365 days	0.008	0.010
arrests	# of own-arrests (drug deals)		0.007	-0.016
networks	# of arrests (property) by 2nd degree neighbors	30 days	0.007	0.009
networks	# of 1st degree neighbors victimized (domestic incident)	730 days	0.007	0.006
arrests	# of own-arrests (resist/obstruct officer)	Total	0.007	0.022

Note: Features are listed in descending order of residualized correlation, except for the first feature. The first column shows the set to which each feature belongs. The second column provides a description of the feature. Text in parentheses indicate a subtype of the feature. The third column shows, where appropriate, the time window over which the feature was measured. Time windows listed as "Total" indicate features that look back to the beginning of our data (August 1999). The fourth and fifth columns show the correlation between the residualized and unresidualized version of the feature and the outcome, respectively.

Table B.7: Top 50 features from the stepwise residualization procedure when not given access to network features

Feature Set	Description	Time Window	Correlations Residualized	Original
arrests	# of own-arrests (any)	730 days	0.109	0.109
arrests	Ever gang-affiliated		0.060	0.095
victims	# of victimizations (shootings)	Total	0.043	0.064
demographics	Age (modal)		-0.034	-0.052
arrests	# of own-arrests (gambling)	Total	0.032	0.068
arrests	# of own-arrests (reckless conduct)	730 days	0.030	0.067
arrests	# of own-arrests (violent)	Total	0.024	0.074
arrests	# of own-arrests (criminal trespass)	270 days	0.020	0.036
arrests	# of own-arrests (solicitation)	730 days	0.019	0.055
demographics	Sex (most recent)		-0.015	-0.017
demographics	Missing date of birth		-0.015	-0.014
arrests	# of own-arrests (gang loitering)	Total	0.015	0.043
arrests	# of own-arrests (drug)	Total	0.014	0.049
victims	# of victimizations (shootings)	270 days	0.014	0.040
demographics	# of unique police beats		-0.013	0.027
arrests	# of own-arrests (robbery)	Total	0.013	0.056
arrests	# of own-arrests (reckless conduct)	Total	0.013	0.072
arrests	# of own-arrests (public alcohol consumption)	Total	-0.011	0.010
arrests	# of own-arrests (firearm possession)	Total	0.011	0.035
arrests	# of own-arrests (any)	365 days	0.011	0.095
arrests	# of own-arrests (drug deals)		0.011	-0.016
arrests	# of own-arrests (burglary)	Total	0.011	0.036
arrests	# of own-arrests (drug)	730 days	0.010	0.040
arrests	# of own-arrests (replica firearm)	Total	0.010	0.030
arrests	# of own-arrests (property)	365 days	-0.010	0.030
arrests	# of own-arrests (chicago municipal code)	730 days	-0.010	0.041
victims	# of victimizations (gun assault or battery)	Total	0.009	0.046
arrests	# of own-arrests (obstructing identification)	Total	0.009	0.033
arrests	# of own-arrests (fbi code wrt)	270 days	-0.009	0.030
arrests	# of own-arrests (solicitation)	Total	0.009	0.060
arrests	# of own-arrests (gun assault or battery)	Total	0.009	0.034
arrests	# of own-arrests (weapons possession)	Total	0.008	0.031
arrests	# of own-arrests (aggravated assault school employee)	Total	0.008	0.027
arrests	# of own-arrests (drug paraphenelia possession)	730 days	-0.008	0.001
arrests	# of own-arrests (gang name indicator)		-0.008	-0.014
arrests	# of own-arrests (domestic)	365 days	-0.008	0.010
arrests	# of own-arrests (disorderly conduct)	Total	0.008	0.024
arrests	# of own-arrests (heroin possesion)	Total	-0.008	0.008
arrests	# of own-arrests (resist/obstruct officer)	Total	0.007	0.022
arrests	# of own-arrests (aggravated robbery)	Total	0.007	0.028
arrests	# of own-arrests (gun assault or battery)	365 days	0.007	0.016
arrests	# of own-arrests (gun battery)	730 days	-0.007	0.011
arrests	# of own-arrests (manufacturing/dealing public school)	Total	0.007	0.027
arrests	# of own-arrests (possession controlled substance)	365 days	-0.007	0.019
arrests	# of own-arrests (robbery)	Total	-0.007	0.004
arrests	# of own-arrests (public alcohol consumption)	730 days	-0.007	0.013
arrests	# of own-arrests (aggravated battery)	Total	0.007	0.019
arrests	# of own-arrests (stolen property)	365 days	-0.007	-0.002
arrests	# of own-arrests (probation violation)	Total	0.006	0.022
arrests	# of own-arrests (violent)	270 days	-0.006	0.027

Note: See bottom of Table B.6 for column definitions.

Table B.8: Top 50 features from the stepwise residualization procedure when not given access to own-arrest features

Feature Set	Description	Time Window	Correlations Residualized	Original
networks	# of 1st degree neighbors ever gang-affiliated	365 days	0.120	0.120
victims	# of victimizations (shootings)	Total	0.044	0.064
networks	# of 1st degree neighbors arrested (drug)	Total	0.039	0.114
demographics	Age (modal)		-0.032	-0.052
demographics	Missing date of birth		-0.023	-0.014
networks	# of 2nd degree neighbors arrested (drug)	30 days	-0.020	0.025
victims	# of victimizations (gun battery)	Total	0.017	0.057
networks	# of arrests (violent) by 1st degree neighbors	Total	0.016	0.089
demographics	Sex (most recent)		-0.016	-0.017
networks	# of arrests (drug) by 2nd degree neighbors	30 days	0.015	0.032
victims	# of victimizations (shootings)	270 days	0.013	0.040
networks	# of 1st degree neighbors victimized (domestic incident)	Total	-0.013	0.018
networks	# of 1st degree neighbors arrested (property)	Total	0.012	0.105
networks	# of 1st degree neighbors arrested (drug)	730 days	0.012	0.102
networks	# of 1st degree neighbors arrested (drug deals)	Total	0.012	0.077
networks	# of 2nd degree neighbors victimized (gun assault or battery)	60 days	0.012	0.013
networks	# of victimizations (gun battery) by 2nd degree neighbors	60 days	0.012	0.006
networks	# of 2nd degree neighbors		0.012	0.075
networks	# of arrests (any) by 1st degree neighbors	Total	-0.011	0.097
victims	# of victimizations (days since first victimization)		-0.011	-0.020
networks	# of arrests (any) by 1st degree neighbors	365 days	0.011	0.102
networks	# of arrests (gun battery) by 2nd degree neighbors	730 days	-0.011	0.006
networks	# of arrests (violent) by 2nd degree neighbors	270 days	-0.011	0.028
networks	# of 2nd degree neighbors victimized (gun battery)	60 days	-0.010	0.005
networks	# of 1st degree neighbors arrested (gun robberies)	Total	0.010	0.044
networks	# of arrests (any) by 1st degree neighbors	60 days	-0.010	0.062
networks	# of 2nd degree neighbors ever gang-affiliated	730 days	0.009	0.090
networks	# of 2nd degree neighbors arrested (any)	730 days	-0.009	0.090
networks	# of arrests (gun robberies) by 1st degree neighbors	Total	-0.009	0.038
networks	# of 2nd degree neighbors		-0.009	0.016
networks	# of victimizations (shootings) by 1st degree neighbors		0.008	0.033
networks	# of arrests (drug) by 2nd degree neighbors	730 days	-0.008	0.057
networks	# of 1st degree neighbors arrested (gun assault or battery)	Total	0.007	0.035
networks	# of arrests (gun robberies) by 1st degree neighbors	730 days	-0.007	0.018
networks	# of arrests (property) by 2nd degree neighbors	30 days	-0.007	0.009
demographics	# of unique police beats		0.007	0.027
networks	# of 2nd degree neighbors arrested (drug)	60 days	0.007	0.034
networks	# of arrests (any) by 1st degree neighbors	90 days	0.007	0.074
victims	# of victimizations (any victimization)	730 days	-0.007	-0.014
networks	# of arrests (drug deals) by 2nd degree neighbors	90 days	-0.007	0.014
networks	# of 1st degree neighbors arrested (domestic)	60 days	0.006	0.012
networks	# of arrests (domestic) by 1st degree neighbors	Total	-0.006	0.033
networks	# of arrests (drug deals) by 1st degree neighbors		0.006	0.036
networks	# of arrests (property) by 2nd degree neighbors	Total	-0.006	0.055
networks	# of arrests (domestic) by 1st degree neighbors	365 days	-0.006	0.022
networks	# of arrests (domestic) by 1st degree neighbors	30 days	-0.006	0.004
demographics	Number of unique races recorded		-0.006	-0.006
networks	# of 1st degree neighbors arrested (domestic)	730 days	0.006	0.035
networks	# of 1st degree neighbors victimized (shootings)	365 days	-0.006	0.038
networks	# of 1st degree neighbors arrested (any)	30 days	0.006	0.060

Note: See bottom of Table B.6 for column definitions.

Table B.9: Top 50 features from the stepwise residualization procedure when not given access to network or own-arrest features

Feature Set	Description	Time Window	Correlations Residualized	Original
victims	# of victimizations (shootings)	Total	0.064	0.064
demographics	Age (modal)		-0.048	-0.052
demographics	Missing date of birth		-0.036	-0.014
demographics	# of unique police beats		0.027	0.027
victims	# of victimizations (any victimization)	Total	-0.023	-0.017
victims	# of victimizations (gun battery)	Total	0.019	0.057
demographics	Sex (most recent)		-0.017	-0.017
victims	# of victimizations (shootings)	270 days	0.015	0.040
victims	# of victimizations (days since first domestic incident)		-0.013	-0.007
victims	# of victimizations (property)		-0.011	-0.007
victims	# of victimizations (gun assault or battery)	730 days	0.011	0.044
victims	# of victimizations (any victimization)	730 days	-0.007	-0.014
victims	# of victimizations (gun assault or battery)	90 days	0.007	0.022
victims	# of victimizations (aggravated battery)	Total	0.007	0.035
demographics	Number of unique races recorded		-0.006	-0.006
victims	# of victimizations (simple domestic battery)	Total	-0.006	-0.014
victims	# of victimizations (property)	Total	-0.006	-0.031
victims	# of victimizations (to property)	Total	0.006	-0.017
victims	# of victimizations (gun battery)	60 days	-0.006	0.013
victims	# of victimizations (gun assault or battery)		0.006	-0.039
victims	# of victimizations (days since first victimization)		-0.006	-0.020
victims	# of victimizations (gun battery)	270 days	-0.005	0.037
victims	# of victimizations (child abduction)	Total	0.005	0.007
victims	# of victimizations (days since last victimization)		0.005	0.006
victims	# of victimizations (aggravated battery)	730 days	0.005	0.033
victims	# of victimizations (criminal sexual assault)	Total	-0.004	-0.005
victims	# of victimizations (property)		-0.004	0.002
victims	# of victimizations (days since first shooting victimization)		-0.004	-0.028
victims	# of victimizations (shootings)	365 days	-0.004	0.041
demographics	Modal race indicator		-0.004	-0.003
victims	# of victimizations (aggravated handgun)	Total	0.003	0.046
victims	# of victimizations (gun battery)		0.003	-0.037
victims	# of victimizations (gun assault or battery)		0.003	-0.029
victims	# of victimizations (child endangerment)	Total	0.003	0.010
victims	# of victimizations (gun robberies)		0.003	-0.004
victims	# of victimizations (aggravated domestic battery)	Total	0.003	0.007
victims	# of victimizations (gun robberies)	60 days	0.003	0.004
victims	# of victimizations (attempted strongram)	Total	-0.003	-0.003
victims	# of victimizations (telephone threat)	Total	-0.003	-0.011
victims	# of victimizations (armed knife)	Total	-0.003	-0.003
victims	# of victimizations (gun assault or battery)	Total	-0.003	0.046
demographics	Modal police beat		0.003	0.001
victims	# of victimizations (gun robberies)		-0.003	-0.009
victims	# of victimizations (shootings)		0.003	0.034
victims	# of victimizations (aggravated)	Total	-0.003	-0.003
victims	# of victimizations (gun battery)		-0.003	-0.030
victims	# of victimizations (gun battery)	90 days	-0.003	0.020
victims	# of victimizations (gun battery)	365 days	0.003	0.039
victims	# of victimizations (aggravated (other dangerous weapon))	Total	0.002	0.009
victims	# of victimizations (retail theft)	Total	0.002	-0.002

Note: See bottom of Table B.6 for column definitions.

Table B.10: Predictive performance for limited feature sets chosen by the stepwise residualization procedure

	Top 500			Top 3,381		
Feature Set	Precision	Recall	Total Recall	Precision	Recall	Total Recall
Full	0.130	0.029	0.019	0.087	0.130	0.087
Full - Top 50	0.118	0.026	0.017	0.089	0.133	0.089
No Own Arrests - Top 50	0.120	0.027	0.018	0.088	0.132	0.088
No Networks - Top 50	0.114	0.025	0.017	0.085	0.127	0.085
No Own Arrests or Networks - Top 50	0.066	0.015	0.010	0.060	0.090	0.060

Note: Performance and recall from models trained to predict shooting victimization during the 18-month outcome period starting April 1, 2018. Models differ based on the feature sets available to them during training. Model performance is evaluated on shooting victimization during the outcome period, for the $k = 500$ and $k = 3,381$ people with the highest predicted risk of shooting victimization.