A novel machine-learning tool to identify community risk for firearm violence: The Firearm Violence Vulnerability Index

Ann M. Polcari, MD, MPH, MSGH, Lea E. Hoefer, MD, Tanya L. Zakrison, MD, MPH, Jennifer T. Cone, MD, MS, Marion C. W. Henry, MD, MPH, Selwyn O. Rogers, MD, MPH, Mark B. Slidell, MD, MPH, *and* **Andrew J. Benjamin, MD, MS**, *Chicago, Illinois*

BACKGROUND: Firearm violence in the United States is a public health crisis, but accessing accurate firearm assault data to inform prevention strategies is a challenge. Vulnerability indices have been used in other fields to better characterize and identify at-risk populations during crises, but no tool currently exists to predict where rates of firearm violence are highest. We sought to develop and validate a novel machine-learning algorithm, the Firearm Violence Vulnerability Index (FVVI), to forecast community risk for shooting incidents, fill data gaps, and enhance prevention efforts.

METHODS: Open-access 2015 to 2022 fatal and nonfatal shooting incident data from Baltimore, Boston, Chicago, Cincinnati, Los Angeles, New York City, Philadelphia, and Rochester were merged on census tract with 30 population characteristics derived from the 2020 American Community Survey. The data set was split into training (80%) and validation (20%) sets; Chicago data were withheld for an unseen test set. XGBoost, a decision tree-based machine-learning algorithm, was used to construct the FVVI model, which predicts shooting incident rates within urban census tracts.

RESULTS: A total of 64,909 shooting incidents in 3,962 census tracts were used to build the model; 14,898 shooting incidents in 766 census tracts were in the test set. Historical third grade math scores and having a parent jailed during childhood were population characteristics exhibiting the greatest impact on FVVI's decision making. The model had strong predictive power in the test set, with a goodness of fit (D^2) of 0.77.

CONCLUSION: The Firearm Violence Vulnerability Index accurately predicts firearm violence in urban communities at a granular geographic level based solely on population characteristics. The Firearm Violence Vulnerability Index can fill gaps in currently available firearm violence data while helping to geographically target and identify social or environmental areas of focus for prevention programs. Dissemination of this standardized risk tool could also enhance firearm violence research and resource allocation. (*J Trauma Acute Care Surg.* 2023;95: 128–136. Copyright © 2023 Wolters Kluwer Health, Inc. All rights reserved.)

LEVEL OF EVIDENCE: Prognostic and Epidemiological; Level IV.

KEY WORDS: Injury prevention; firearm violence; vulnerability; machine learning.

Every day in the United States, an estimated 200 Americans are injured in non–suicide-related shootings, resulting in nearly 20,000 deaths per year.[1,2] Despite a variety of efforts by policymakers, local nonprofits, and national organizations to reduce firearm violence, its volume has continued to increase, most recently exacerbated by the COVID-19 pandemic.[3] Certain policies like the Dickey Amendment have severely limited the ability to conduct robust firearm violence research for nearly two decades.[4,5] Thus, the true magnitude of and underlying factors contributing to the firearm violence crisis remain unknown, and the implementation of targeted prevention strategies is hindered.

Currently, there is no centralized national database collecting real-time firearm-related injury data. The data that do exist are incomplete and have restricted access, a narrow focus, and substantial delay in publishing.[6] Local and state trauma registries, as well as many national trauma databases, are neither widely available nor uniformly provide granular information related to shooting incident location or victim home address. While most firearm assaults are nonfatal, many current resources solely report homicides. In addition, the intent behind nonfatal injuries is frequently misclassified or "unknown."[7] Moreover, the landscape of firearm violence can quickly change in the years it takes to distribute this data. Sources that have attempted to circumvent these issues, such as the Centers for Disease Control and Prevention's National Electronic Injury Surveillance System, the Gun Violence Archive, and the Federal Bureau of Investigation's National Incident Based Reporting System have been criticized for missing records.[8–10] In addition, the data that are available are very difficult to merge. Without an understanding of the true scope of US firearm-related violence, it is difficult to formulate appropriate injury prevention programs. Therefore, innovation in capturing firearm violence data is crucial to develop effective prevention strategies.

Submitted: December 1, 2022, Revised: February 17, 2023, Accepted: March 22, 2023, Published online: April 4, 2023.

From the Department of Surgery (A.M.P., L.E.H.), University of Chicago Medicine; Section of Trauma and Acute Care Surgery, Department of Surgery (T.L.Z., J.T.C., S.O.R., A.J.B.); Section of Pediatric Surgery, Department of Surgery (M.C.W.H.), University of Chicago, Chicago, Illinois; and Division of General Pediatric Surgery, Department of Surgery (M.B.S.), The Johns Hopkins University, Baltimore, Maryland.

This study was presented at the 36th Annual Meeting of the Eastern Association for the Surgery of Trauma, January 17–21, 2023, in Lake Buena Vista, Florida.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.jtrauma.com).

Address for correspondence: Ann M. Polcari, MD, MPH, MSGH, 5481 South Maryland Avenue, MC5030, Chicago, IL 60637; email: ann.polcari@uchicagomedicine.org.

DOI: 10.1097/TA.0000000000003992

A core principle of injury prevention is understanding vulnerable populations at risk. A plethora of analyses have previously sought to describe community risk for firearm violence based on the social determinants of health, but heterogeneity among these studies regarding geographic level of interest and the factors chosen to define risk limits their generalizability and utility in developing scalable prevention programs. More recently, several nationally standardized indices that use social, structural, and geospatial determinants of health to identify populations at greatest risk during public health crises have been published. Examples include the Centers for Disease Control and Prevention's Social Vulnerability Index (SVI), the Neighborhood Atlas Area Deprivation Index (ADI), and the Childhood Opportunity Index (COI) from diversitydatakids.org.[11–13] Notably, several recent studies have used these indices to examine firearm violence, revealing that concentrated geographic areas with high community vulnerability also have high rates of shooting incidents (A.M. Polcari, MD, MPH, MSGH, unpublished data, March 2023).[14] This work implies that certain population characteristics could assist in determining where prevention efforts should be focused even without accessible shooting incident data and that standardization of risk in firearm violence research is possible. Although recently extrapolated to firearm-related injury, these indices were ultimately constructed for other purposes: SVI was made to mitigate loss after natural disasters, ADI was created to understand health care–related outcomes, and COI was established to show inequity in childhood development. To date, no index has been specifically designed to identify populations at the greatest risk for firearm violence or forecast shooting incident rates. The purpose of this study was to develop and validate a novel machine-learning algorithm that predicts firearm violence in the urban United States based solely on social, structural, and geospatial determinants of health, which could augment firearm injury research efforts and prevention planning by filling data gaps, standardizing risk, assisting in trauma system planning, and projecting changes over time.

PATIENTS AND METHODS

Data Sources and Study Cohort

In this study, we developed a machine-learning algorithm to predict rates of firearm violence within urban US census tracts. Our cohort was generated from a convenience sample of eight US cities, Baltimore (BAL), Boston (BOS), Chicago (CHI), Cincinnati (CIN), Los Angeles (LA), New York City (NYC), Philadelphia (PHL), and Rochester (RNY), which provide open-access incident location records for both fatal and nonfatal shootings. Shooting incident data for 2015 to 2021 were obtained from Open Baltimore, Analyze Boston, the Chicago Data Portal, Cincinnati Open Data Portal, County of Los Angeles Open Data, NYC Open Data, the Philadelphia Office of the Controller, and Rochester Police Department Open Data Portal. Firearm assaults and firearm-related homicides in persons of all ages were identified using Unified Crime Reporting codes and were retained in the data set. Self-inflicted and accidental injuries were excluded. Latitude and longitude coordinates for each shooting incident were geocoded into census tract using the publicly available Federal Communications Commission API. Population estimates were extracted from the 2020 American Community Survey (ACS) 5-year estimates.[15] The prevalence of shooting incidents within each census tract during our study's timeframe was then calculated per 1,000 persons. Census tracts with a population less than 1,000 were excluded. Institutional review board review was waived given the use of open-access deidentified data. The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guideline for both model development and validation was used to ensure proper reporting of methods, results, and discussion (Supplemental Digital Content, Supplementary Data 1, http://links.lww.com/TA/C965; Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis checklist).

Predictors and Outcomes

We a priori chose a set of 30 population characteristics to include in our prediction model, which we named the Firearm Violence Vulnerability Index (FVVI) (Fig. 1). These variables were selected based on literature review, established public health principles like the social and structural determinants of health, and the evaluation of various published measures of community deprivation. We abstracted each variable at the census tract level from the Census Bureau's 2020 ACS 5-year estimates and Opportunity Atlas[16] (Supplemental Digital Content, Supplementary Table 1, http://links.lww.com/TA/C966; FVVI Population Characteristic Explanations and Data Sources). To provide an organized framework for features of the FVVI, we grouped the variables into seven domains: Population Demographics, Household Composition, Socioeconomic Status, Education, Housing Characteristics, Access to Basic Needs, and Employment Opportunity. The model's primary outcome is shooting rate per 1,000 people within a census tract.

Model Development

Firearm violence data from BAL, BOS, CIN, LA, NYC, PHL, and RNY were merged on census tract with each variable selected for the FVVI; data from CHI were withheld to independently assess the final model's accuracy. This cohort was then randomly split into training (80%) and validation (20%) sets to build the model.

We used a decision tree–based machine-learning algorithm called Extreme Gradient Boosting (XGBoost) to construct the FVVI model.[17] In short, XGBoost uses the chosen input features and training data to create an initial set of decision trees in parallel, followed by multiple iterations of new decision trees that improve upon any errors identified in the prior set. Each iteration is cross validated using the designated validation data until the ideal number of iterations is reached. The final output model is a weighted sum of all the decision trees that were ultimately generated, which lowers the risk of overfitting when compared with a single, more complex decision tree.[18] We optimized the model hyperparameters using Optuna, a Bayesian automated hyperparameter optimization framework. We also elected to fit the FVVI to a Poisson regression learning objective based on the nonnormal distribution of our data. Census tracts with missing population characteristic data were not imputed because of XGBoost's ability to inherently handle missingness. When XGBoost encounters missing data for an input variable during training, it uses Sparsity-aware Split Finding. Essentially,

Firearm Violence Vulnerability Index (FVVI)	Population Demographics	Male <18 years old ≥65 years or Older Speaks English Less than Well
	Household Composition	Single Parent Household Parent Incarcerated in Childhood Grandparent Primary Caregiver Civilian with a Disability
	Socioeconomic Status	Income Percentile Below Poverty Unemployment Rate Public Cash Assistance Food Stamps/SNAP
	Education	≥3 years old Enrolled in School Historical 3rd Grade Math Scores ≤8th Grade Education High School Degree Bachelor's Degree
	Housing Characteristics	Owner-Occupied Housing Rent ≥35% of income ≥20 Unit Buildings Vacant Housing Units Crowding
	Access to Basic Needs	No Health Insurance Lacks complete plumbing Broadband Internet Subscription No Vehicle
	Employment Opportunity	Available jobs Job Growth Long Commute

Figure 1. The FVVI. The FVVI is made of 30 population characteristics derived from the US Census Bureau's ACS and Opportunity Atlas. These population characteristics were grouped into seven domains. The FVVI machine-learning model predicts shooting incident rates within a census tract.

multiple threshold values for the missing variable are tested, and the value with the greatest predictive benefit is selected. This value is then set as default for all missing data within the variable at that decision point. Following optimization, the model was retrained on the entire cohort before testing. Of note, the output of the model, that is, shootings per 1,000 persons in a census tract, was rescaled by standardizing all values between 0 and 1 to ease within- and between-city comparisons after future expansion to a broader geographic scale and for studying series of time.

To explain how each population characteristic effected the final model, we generated two representations of feature importance: accuracy gain and permutation. Accuracy gain feature importance tells the relative contribution of individual variables to the final model, based on the variable's role in each decision tree encompassed by the final model. Permutation feature importance reveals the variables with the highest impact in assigning shooting incident risk for the training and validation sets combined. In the calculation of permutation feature importance, a variable is considered more important if its removal from the model results in a significant increase in error.

Model Assessment

We assessed the FVVI's ability to predict shooting incidents within a census tract using unseen test data from CHI. Firearm Violence Vulnerability Index model performance was evaluated by calculating the mean Poisson deviance and deviance goodness of fit (D^2). The mean Poisson deviance is a measure of how well the model fits the observed data sample, similar to the root mean squared error of a linear model. The deviance goodness of fit explains the degree of this variance that is accounted for by the model, akin to r^2 in a linear model.

We then used the Shapley Additive Explanations (SHAP) method to demonstrate how each population characteristic affected the FVVI's decision making regarding shooting incidents in CHI. Shapley Additive Explanations values add interpretability to complex algorithms such as XGBoost, as they represent the log odds of a variable's contribution to each individual prediction made by the model. The summation of SHAP values demonstrates a model's overall behavior when making predictions in a given cohort. In addition, choropleth maps were created for both true shooting incident rates in CHI during our study's timeframe, as well as those predicted by the FVVI.

Statistical Analysis

All statistical analyses were conducted in Python 3.9 (Python Software Foundation, www.python.org). Characteristics of the training, validation, and test sets were compared using Wilcoxon rank-sum tests in Statsmodels v0.13.2. The FVVI model was developed using sklearn version 1.1.1, xgboost version 1.5.0, and Optuna version 3.0.3. Choropleth maps were created using geopandas version 0.9.1.

RESULTS

Study Cohort

A total of 64,909 shooting incidents in 3,962 census tracts were used to develop the model (BAL, BOS, CIN, LA, NYC, PHL, RNY), of which 175 (5.9%) were missing data for 1 or more FVVI population characteristics. A total of 3,169 census tracts were used in the training set and 793 in the validation set. In the test set (CHI), there were 14,898 shootings incidents in 776 census tracts. The population characteristics of the combined training and validation set were generally different from the test set (Table 1). Overall, data for 4.3% of the US population in 2020 were included in the development of the FVVI.

Model Characteristics

The population characteristics exhibiting the greatest influence on FVVI's prediction for shooting incidents per 1,000 people in a census tract are shown in Figure 2. In both accuracy gain and permutation feature importance, historical third-grade math scores of adults approximately 30 years of age during our study's timeframe were the most important feature in the model, followed by this same population's percentage of parents incarcerated during their childhood. The feature with the third greatest importance in the FVVI model by both accuracy gain and permutation importance was a measure of the built environment: vacant housing units. This was followed by food assistance programs in the accuracy gain ranking and income percentile in the permutation ranking. Eight of the top 10 population variables of importance in both the accuracy gain and permutation rankings were the same. No FVVI domain was clearly dominant based on either method of ranking feature importance.

Model Performance

We evaluated the FVVI's performance using data from CHI that was unseen by the model during training. The FVVI had a high predictive ability for shooting incidents per 1,000 people in a census tract, with a mean Poisson deviance of 2.29 and overall goodness of fit as calculated by D^2 of 0.77. Table 2 demonstrates FVVI predicted and actual shooting rates from 2015 to 2021. An FVVI value of 0.7 or higher appears to confer a notable increase in the risk of shooting incidents within a census tract. A geospatial comparison of actual shooting incidents per 1,000 persons in CHI census tracts versus those predicted by the FVVI is shown in Figure 3.

To demonstrate the way in which population characteristics were used by the FVVI model to predict shootings within CHI, we used SHAP summary plots (Fig. 4). Shapley Additive

TABLE 1. Combined Training and Validation Versus Test Cohort Comparison

FVVI Population Characteristic (% Within a Census Tract, Unless Otherwise Specified)	Combined Training and Validation Set (n = 2,967 Census Tracts) Mean (SD)	Test Set (n = 697 Census Tracts) Mean (SD)	*p*
Male	47.77 (4.73)	48.08 (5.29)	0.119
<18 y old	20.77 (7.80)	21.10 (7.73)	0.032
≥65 y or older	14.73 (7.01)	13.08 (6.65)	<0.001*
Speaks English less than well	17.86 (15.44)	13.63 (13.82)	<0.001*
Single parent household	10.96 (8.36)	12.04 (8.74)	<0.001*
Parent incarcerated in childhood	1.81 (2.11)	2.77 (2.48)	<0.001*
Grandparent primary caregiver	26.47 (38.19)	24.16 (35.60)	0.037*
Civilian with a disability	12.39 (6.32)	11.96 (6.48)	0.054
Income percentile (percentile)	0.42 (0.31)	0.39 (0.32)	0.003*
Below poverty	14.69 (12.66)	15.79 (12.75)	0.014*
Unemployment rate	7.36 (5.09)	10.48 (8.17)	<0.001*
Public cash assistance	4.96 (4.81)	3.70 (3.29)	<0.001*
Food stamps/SNAP	20.81 (16.33)	22.23 (16.42)	0.024*
≥3 y old enrolled in school	6.88 (5.70)	7.19 (6.01)	0.389
Third-grade math scores (nationally normalized NAEP score)	2.88 (0.65)	2.21 (0.23)	<0.001*
≤8th Grade education	7.88 (6.39)	7.60 (7.70)	<0.001*
High school degree	83.42 (10.59)	84.17 (11.07)	0.039*
Bachelor's degree	35.51 (21.60)	35.12 (25.14)	0.010*
Owner-occupied housing	40.38 (24.38)	44.63 (20.29)	<0.001*
Rent ≥35% of income	44.48 (14.64)	41.92 (16.04)	<0.001*
≥20 Unit buildings	29.90 (31.73)	16.52 (24.89)	<0.001*
Vacant housing units	9.79 (7.35)	12.10 (7.80)	<0.001*
Crowding	2.75 (3.40)	1.24 (1.74)	<0.001*
No health insurance	6.72 (4.53)	9.98 (6.20)	<0.001*
Lacks complete plumbing	0.38 (0.96)	0.42 (1.28)	0.914
Broadband internet subscription	82.93 (10.36)	80.48 (10.49)	<0.001*
No vehicle	42.19 (23.88)	26.18 (15.12)	<0.001*
Available jobs (total no. jobs)	735,721.33 (761,271.36)	385,950.32 (317,790.65)	<0.001*
Job growth	2.56 (7.72)	0.26 (8.95)	<0.001*
Long commute, min	38.98 (8.93)	35.66 (5.81)	<0.001*

*Statistically significant *p* value.
NAEP, National Assessment of Education Progress.

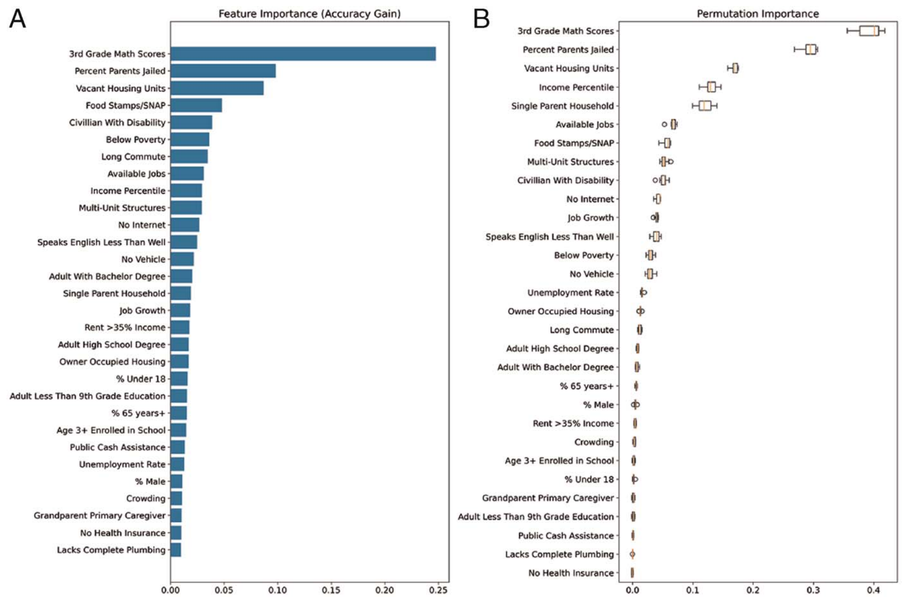

Figure 2. Firearm Violence Vulnerability Index model feature importance. (*A*) Accuracy gain feature importance demonstrating the relative contribution of each population characteristic to the model. (*B*) Permutation feature importance demonstrating the population characteristics that, when removed from the model, result in the greatest increase in prediction error.

Explanations values are ranked by the magnitude of each population characteristic's contribution toward the model's predictions. The distribution of the bee swarm plot reveals the underlying relationships on which these predictions are made. Of note, the longer the bee swarm extension in either direction, the stronger an impact that characteristic has on the model's prediction. The top two population characteristics used to predict shooting incidents in CHI were historical third-grade math scores and percentage of parents jailed for persons around 30 years old at the time of our study. Figure 4*B* shows that low mean third-grade math scores led to a prediction of more shootings, while fewer parents in jail resulted in a prediction of fewer shootings. These were followed by income percentile and single parent households. The SHAP summary demonstrates an inverse relationship between income percentile and predicted shooting incidents. A low number of single-parent households within a census tract had a strong negative effect on shooting incident projections.

TABLE 2. Predicted Versus Actual Shootings Per Census Tract by FVVI Decile in Chicago From 2015 to 2021

FVVI Decile	FVVI Predicted Shootings per 1,000 Persons	Actual Shootings per 1,000 Persons
0.1	1.13	0.54
0.2	1.51	0.94
0.3	1.99	1.48
0.4	2.52	2.21
0.5	3.55	3.36
0.6	5.31	5.94
0.7	8.81	9.28
0.8	14.4	15.15
0.9	20.93	22.48
1.0	58.78	64.80

Other highly impactful relationships in FVVI's predictions in CHI include fewer shootings in census tracts with less people proficient in English or requiring food stamps but more shootings in census tracts with more vacant housing units. Notably, three of four population characteristics in our Housing Characteristic domain were ranked in the top-10 SHAP values for FVVI's prediction algorithm in CHI.

DISCUSSION

In this study, we developed a novel machine-learning algorithm, named the FVVI, to predict firearm violence within urban US census tracts based on social, structural, and environmental determinants of health. Our model was developed using open-access shooting incident data and population characteristics from seven major US cities (BAL, BOS, CIN, LA, NYC, PHL, RNY). It was able to predict shooting incidents very accurately in an independent city (CHI) with a largely statistically different population composition than the training cohort, demonstrating generalizability to other US urban environments. Thus, we believe that the FVVI can potentially inform prevention strategies by predicting which communities are particularly susceptible to firearm violence, even when data on shooting incidents are either missing or unknown. Given the current state of firearm violence in the United States and lack of easily accessible shooting incident data, we think that an innovative tool like the FVVI could prove valuable in developing injury prevention programs for vulnerable, at-risk populations.

To our knowledge, FVVI is the first model created to specifically predict firearm violence at the census tract level based on publicly available population characteristics. Several studies have previously established relationships between the social and structural determinants of health and firearm homicides in children and adults, although there is significant heterogeneity

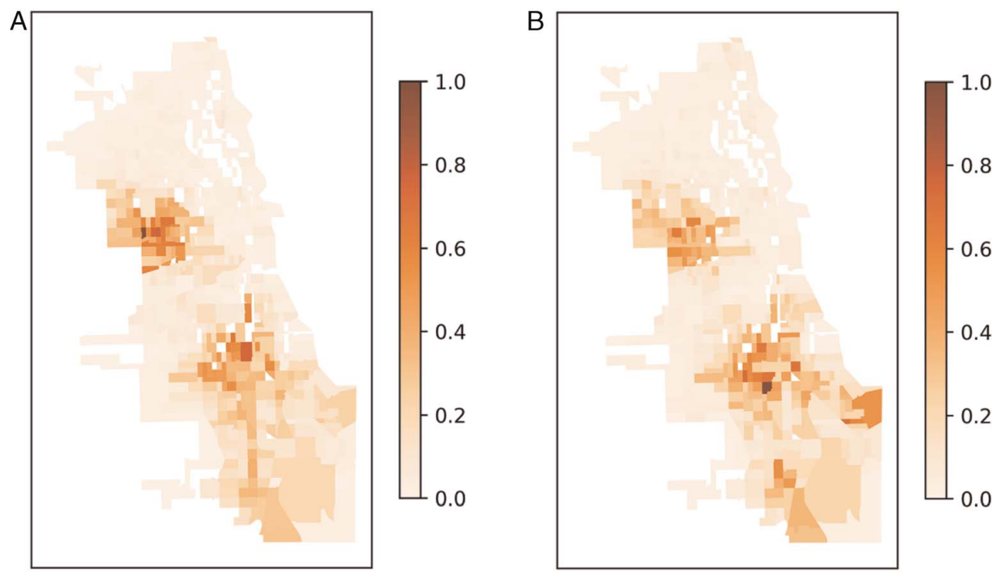

Figure 3. Geospatial representation of model assessment in Chicago. Choropleth maps representing (*A*) actual and (*B*) FVVI predictions for shooting incidents per 1,000 persons within a census tract from 2015 to 2021. All rates were normalized and percentile-ranked between 0 and 1 for interpretability.

among these studies regarding the population characteristics included and the geographic level of interest.[19–23] A series of recent work has attempted to rectify this by using the SVI to identify high-impact social factors and geospatial areas of focus for prevention planning across multiple US cities at the census tract level (Polcari, MD, MPH, MSGH et al., unpublished data,

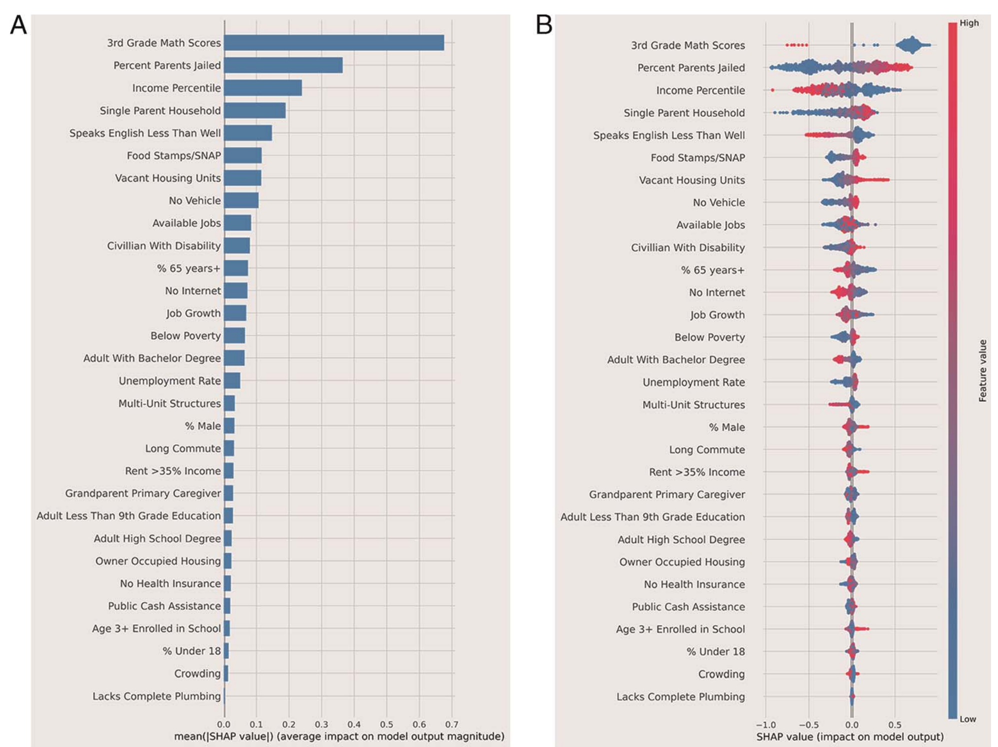

Figure 4. Shapley Additive Explanations values for Chicago test set. Shapley Additive Explanations methods were used to demonstrate how the FVVI model made predictions in Chicago. (*A*) The contribution of each population characteristic to the model's decision making ranked by SHAP value. (*B*) Bee swarm plots demonstrating the association of each population characteristic with shooting incidents as identified by the model. In our study, each point on the SHAP plot corresponds to an individual census tract in CHI. The collective of points for CHI census tracts on the SHAP plot create a bee swarm that represents how the FVVI model is associating a population characteristic with shooting incidents.

March 2023).[14] Other groups have also demonstrated relationships between the ADI, COI, the Gini Index, and firearm violence using local trauma registry or county-level homicide data.[24–27] This collective of research inspired the FVVI and informed many of the population characteristics selected for our algorithm. However, there are several key differences between this prior work and the FVVI, from data sources to statistical methods.

The FVVI was trained and assessed using ACS population characteristics and publicly available shooting incident data at the census tract level. We thought that it was important to include both fatal and nonfatal firearm assaults. Nonfatal firearm injuries are not only more frequent but also have profound effects on individuals and communities and therefore are important to understand when developing prevention programs. To build a robust machine-learning model, we needed to gather a sizable array of these data. Nonetheless, obtaining accurate data for these incidents, especially associated with geolocation, are a challenge.[6] Since there is no real-time, useable, open-access national data set containing this granular information, we convenience sampled a group of cities with firearm violence data published online. These data are most often derived from police records, where standard Unified Crime Reporting codes are used to specifically categorize assaults with a firearm. There are few studies that use this type of data to study firearm violence, but there are certain benefits. The primary advantage is the frequent association of geolocation with a shooting incident, which allows for geospatial analyses not typically available in databases that rely on medical records. Police data are also released more rapidly than hospital-based data, sometimes on a weekly or monthly basis. Moreover, these data include victims who are likely unaccounted for by hospital records, such as those who are deemed deceased at the scene.[9] Currently, there is no mandate requiring police departments to openly report detailed shooting incidents to the public or the to the Federal Bureau of Investigation's National Incident Based Reporting System, an unfortunate reality for firearm violence researchers and public health practitioners. Nevertheless, the power of the FVVI is in filling these gaps. As our independent assessment of CHI demonstrates, FVVI has the capacity to accurately predict urban shooting incidents at a granular level based exclusively on nationally available, annually recurring, open-access population characteristic data.

The Firearm Violence Vulnerability Index's ability to accurately forecast where shooting incidents are most likely to occur lies in the methodology. We chose to develop FVVI using machine-learning techniques to create a highly discriminative model. Because machine-learning leverages complex algorithms and advanced computing power, it can detect nonlinear patterns in large data sets with high dimensionality that are too elaborate for standard statistical methods.[17] Another benefit to the machine-learning technique is that the model can continue to "learn" as new data become available, enhancing its predictive power over time. As such, our model could be used to assess proposed interventions that aim to affect its input variables, thereby predicting how firearm violence might change after implementation.

We believe that the FVVI could be a powerful and practical tool for use in firearm violence prevention. Among other criteria, successful prevention programs need to be targeted to the appropriate population and address local and systemic needs. As mentioned, the FVVI can assist in targeting populations that have the highest rates of firearm violence at a granular geographic level, even if shooting incident data are unavailable. It might also allow us to focus prevention efforts on the factors that most increase firearm violence in a community through feature importance analysis and SHAP scores, which can be obtained for any location analyzed. In our CHI test set, low historical third-grade math scores, a history of a parent jailed during childhood, and low-income percentile had the greatest effect on the algorithm's prediction for high rates of shootings. In theory, prevention efforts aimed at these factors, such as policies for equitable school funding and educational resources, social support teams for families with an incarcerated parent, or programs that increase economic opportunity, could help to reduce firearm violence over time. Furthermore, the FVVI could help predict outcomes of these interventions. By manipulating the input population characteristics that a particular prevention approach attempts to change, we can anticipate its potential effect on firearm violence.

Other potential applications of the FVVI are at the broader trauma systems level, as well as the individual trauma patient level. Areas with the highest predicted rates of penetrating trauma should be considered when opening new trauma centers and planning transport systems. At the "micro" level, FVVI could be used by hospital-based violence intervention and prevention specialists to understand the unique needs of trauma patients who are victims of firearm injury, ensuring connections are made to local resources that address their social, structural, and even geospatial determinants of health, before discharge.

Our ultimate goal is to make FVVI openly available for use by public health experts and researchers, similar to SVI, ADI, and COI. We believe that the FVVI could improve the way in which firearm violence prevention programming and research are approached by standardizing risk and providing a basis for comparison across urban settings. It can give nuanced information on a community's vulnerability to firearm violence, provide insight on the population characteristics that contribute most, and offer a way to measure the impact of interventions over time. With even more data and dissemination, we believe that the FVVI could be a valuable tool in firearm violence prevention.

LIMITATIONS

This study has several limitations. First, the aforementioned paucity of publicly available shooting incident data limited the number of cities on which the FVVI model could be trained. With more data, the model's goodness of fit would likely further improve. Likewise, while FVVI's prediction in CHI provides evidence of generalizability to urban settings, the model has not been trained on data from or tested in rural census tracts and cannot reliably be used in this setting at this point. The cities studied were all medium to large in population, so the FVVI's accuracy in smaller cities remains to be seen. Moreover, although machine-learning algorithms can provide superior predictions, the method has been criticized for the "black box" nature of its outputs. However, our use of the XGBoost decision tree-based model allowed us to generate feature importance, and emerging techniques like SHAP values provide additional interpretability.

Most prior studies that geospatially analyze firearm violence use an individual's home address derived from hospital data. In contrast, police data provide an approximate geographic location where a shooting incident occurred. Therefore, the FVVI at a person's home could be different than where he or she suffered a firearm injury. However, not only do many firearm assaults occur away from the home but prior research suggests that the built environment can play a significant role in where and why a shooting occurs.[28–30] This is reflected in our own FVVI feature importance and SHAP values, as the concentration of vacant housing units played a significant role in the algorithm's decision making. Using aspects of the built environment in future iterations of this model might further enhance its predictability.

Finally, our model does not include the social construct of “race” as a population characteristic, which has been incorporated into other published indices. Racial disparities in health outcomes and violence are frequently studied; however, we believe that these associations are largely quantifying the effects of racism and structural injustices in the United States. We also believe that this is represented in the FVVI, as policies like redlining and enduring community disinvestment are reflected in the social, structural, and geospatial factors that most contribute to firearm violence vulnerability.

CONCLUSION

Given the continued rise in US firearm violence, innovative methods to inform injury prevention strategies are critical. The FVVI is a novel machine-learning algorithm that can predict the rate of shooting incidents within urban census tracts based on open-access population characteristics, thus filling gaps in currently available firearm violence data. The Firearm Violence Vulnerability Index predictions can not only assist in targeting interventions to the most vulnerable communities geographically but can also identify social, structural, and environmental factors on which to focus these prevention efforts. Furthermore, this evidence-derived, standardized framework for firearm violence risk might offer a more systematic approach to research and the distribution of funds in this field. While the FVVI could become more powerful and refined with the addition of new data over time, we believe that it can still urgently enhance ongoing efforts to decrease firearm violence.

AUTHORSHIP

A.M.P. conceptualized and designed the study, interpreted the data, and drafted the article. A.J.B. conceptualized and designed the study, performed the analysis, assisted in interpretation of the data, and performed critical revision of the article. L.E.H., T.L.Z., J.T.C., M.C.W.H., S.O.R., and M.B.S. all contributed to the conception and design of the study; they also performed critical revision of the article. All authors approved the final article as submitted.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Everytown Research & Policy. A more complete picture: the contours of gun injury in the United States. Revised November 2, 2021. Effective October 4, 2020. Available at: https://everytownresearch.org/report/nonfatals-in-the-us/. Accessed November 20, 2022.
2. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. Web-based Inquiry Statistics Query and Reporting System (WISQARS). Centers for Disease Control and Prevention. Atlanta, GA. Available at: www.cdc.gov/ncipc/wisqars. Accessed November 20, 2022.
3. Ssentongo P, Fronterre C, Ssentongo AE, Advani S, Heilbrunn ES, Hazelton JP, et al. Gun violence incidence during the COVID-19 pandemic is higher than before the pandemic in the United States. *Sci Rep*. 2021;11(1):20654.
4. Rostron A. The Dickey Amendment on federal funding for research on gun violence: a legal dissection. *Am J Public Health*. 2018;108(7):865–867.
5. Stark DE, Shah NH. Funding and publication of research on gun violence and other leading causes of death. *JAMA*. 2017;317(1):84–85.
6. Kaufman EJ, Delgado MK. The epidemiology of firearm injuries in the US: the need for comprehensive, real-time, actionable data. *JAMA*. 2022; 328(12):1177–1178.
7. NORC at The University of Chicago. A Blueprint for U.S. Firearms Data Infrastructure. Effective October 14, 2020. Available at: https://www.norc.org/PDFs/Firearm%20Data%20Infrastructure%20Expert%20Panel/A%20Blueprint%20for%20a%20U.S.%20Firearms%20Data%20Infrastructure_NORC%20Expert%20Panel%20Final%20Report_October%202020.pdf. Accessed November 20, 2022.
8. Kaufman EJ, Passman JE, Jacoby SF, Holena DN, Seamon MJ, MacMillan J, et al. Making the news: victim characteristics associated with media reporting on firearm injury. *Prev Med*. 2020;141:106275.
9. Barber C, Cook PJ, Parker ST. The emerging infrastructure of US firearms injury data. *Prev Med*. 2022;165:107129.
10. Campell S, Nass, D, Nguyen, M. The CDC Says Gun Injuries Are on the Rise. But There Are Big Problems With Its Data. The Trace. October 4, 2018. Available at: https://www.thetrace.org/2018/10/cdc-nonfatal-gun-injury-data-estimate-problems/. Accessed on November 20, 2022.
11. University of Wisconsin School of Medicine and Public Health. Neighborhood Atlas. Maison, WI. Available at: https://www.neighborhoodatlas.medicine.wisc.edu/. Accessed November 1, 2022.
12. diversitydatakids.org. Childhood Opportunity Index 2.0. Available at: https://data.diversitydatakids.org/dataset/coi20-child-opportunity-index-2-0-database. Accessed November 1, 2022.
13. Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry/Geospatial Research A, and Services Program. CDC/ASTDR Social Vulnerability Index. Available at: https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html. Accessed October 1, 2021.
14. Polcari AM, Hoefer LE, Callier K, Zakrison TL, Rogers SO, Henry M, Slidell MB, Benjamin AJ. Social Vulnerability Index is Strongly Associated with Urban Pediatric Firearm Violence: An Analysis of Five Major U.S. Cities. *J Trauma Acute Care Surg*. 2023. doi: 10.1097/TA.0000000000003896.
15. American Community Survey 5-year Data Profiles. United States Census Bureau. Washington, DC. Available at: https://data.census.gov/table?d=ACS+5-Year+Estimates+Data+Profiles. Accessed October 1, 2021.
16. Opportunity Atlas Data Tool. United States Census Bureau. Washington, DC. Available at: https://www.census.gov/programs-surveys/ces/data/public-use-data/opportunity-atlas-data-tables.html. Accessed June 1, 2022.
17. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd SCM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, August 13–17, 2016, New York, NY: Association for Computing Machinery; 2016. 785–794.
18. Salati M, Migliorelli L, Moccia S, Andolfi M, Roncon A, Guiducci GM, et al. A machine learning approach for postoperative outcome prediction: surgical data science application in a thoracic surgery setting. *World J Surg*. 2021;45(5):1585–1594.
19. Trinidad S, Vancil A, Brokamp C, Moody S, Gardner D, Parsons AA, et al. Relationships between socioeconomic deprivation and pediatric firearm-related injury at the neighborhood level. *J Trauma Acute Care Surg*. 2022;93(3):283–290.
20. Stevens J, Leonard J, Reppucci ML, Schroeppel T, Bensard D, Haasz M. Individual and neighborhood level characteristics of pediatric firearm injuries presenting at trauma centers in Colorado. *J Trauma Acute Care Surg*. 2022;93(3):385–393.
21. Kim D. Social determinants of health in relation to firearm-related homicides in the United States: a nationwide multilevel cross-sectional study. *PLoS Med*. 2019;16(12):e1002978.
22. Houghton A, Jackson-Weaver O, Toraih E, Burley N, Byrne T, McGrew P, et al. Firearm homicide mortality is influenced by structural racism in US metropolitan areas. *J Trauma Acute Care Surg*. 2021;91(1):64–71.

23. Formica MK. An eye on disparities, health equity, and racism-the case of firearm injuries in urban youth in the United States and globally. *Pediatr Clin North Am*. 2021;68(2):389–399.
24. Kwon EG, Wang BK, Iverson KR, O'Connell KM, Nehra D, Rice-Townsend SE. Interpersonal violence affecting the pediatric population: patterns of injury and recidivism. *J Pediatr Surg*. 2023;58:136–141.
25. Gastineau KAB, Williams DJ, Hall M, Goyal MK, Wells J, Freundlich KL, et al. Pediatric firearm-related hospital encounters during the SARS-CoV-2 pandemic. *Pediatrics*. 2021;148(2):e2021050223.
26. Abaza R, Lukens-Bull K, Bayouth L, Smotherman C, Tepas J, Crandall M. Gunshot wound incidence as a persistent, tragic symptom of area deprivation. *Surgery*. 2020;168(4):671–675.
27. Rowhani-Rahbar A, Quistberg DA, Morgan ER, Hajat A, Rivara FP. Income inequality and firearm homicide in the US: a county-level cohort study. *Inj Prev*. 2019;25(Suppl 1):i25–i30.
28. Lipton R, Yang X, Braga AA, Goldstick J, Newton M, Rura M. The geography of violence, alcohol outlets, and drug arrests in Boston. *Am J Public Health*. 2013;103(4):657–664.
29. MacDonald J. Community design and crime: the impact of housing and the built environment. *Crime Justice*. 2015;44:333–383.
30. University of Chicago Crime Lab New York. The Impact of Street Lighting on Crime in New York City Public Housing. Effective October 2018. Available at: http://www.modresvetlo.cz/PDF/The_Impact_of_Street_Lighting_on_Crime_in_New_York_City_Public_Housing.pdf. Accessed September 20, 2022.