

Fatal US School Shootings Are More Likely in Targeted and Multi-Victim Events: Evidence from Logistic Regression on Contextual Factors

Amanda Ng

2025-10-25

Abstract

1. Introduction

Gun violence in schools remains one of the most distressing forms of firearm-related harm in the United States (citation). Each incident causes immediate loss and trauma while also increasing public concern over the safety of educational environments. Despite ongoing research, relatively little is known about why some school shootings result in fatalities while others do not. Understanding these factors is important for informing prevention strategies and school safety policies. Previous studies have examined individual-level factors such as shooter motivation, access to firearms, and mental health (citation). Fewer studies have examined contextual-level characteristics—features of the incident and school environment that may affect whether a shooting results in a fatality. This study examines these contextual factors to identify conditions that are associated with higher likelihoods of fatal outcomes.

The analysis draws on The Washington Post School Shootings Database, which documents every act of gunfire on U.S. primary and secondary school campuses during school hours since 1999. The estimand of interest is the probability that a school shooting results in at least one fatality. Using logistic regression, we modeled this probability as a function of several contextual characteristics, including whether the incident involved any injuries, the shooting type, the shooter's demographics and relationship to the school community, and the broader racial and socioeconomic composition of the school.

Results show that shootings involving at least one injury, those classified as targeted attacks, and those committed out by shooters with no apparent or direct relationship to the school were more likely to result in fatalities. Shootings by individuals with no direct relationship to the school appear less predictable and may involve higher intent or premeditation. As such incidents are outside the reach of internal monitoring systems focused on students and staff, these findings emphasize the importance of external security measures, including controlled access to school facilities, improved information sharing with law enforcement and between schools, and preparedness for emergency response. Examining these patterns contributes to understanding the factors associated with school shooting fatality and can support the development of measures to reduce harm to students and school communities.

Section 2: Data describes the data source, alongside the cleaning and preparation of the dataset for analysis. A summary of key variables and exploratory data visualizations is also presented. Section 3: Modelling explains the logistic regression approach used to analyze the relationship between fatality likelihood and contextual characteristics, along with justifications for the model design. Section 4: Results & Discussion presents the results and interprets the effects of predictors, as well as limitations of the study, including data quality and practical applicability of the models. Section 5: Conclusion summarizes the main findings and outlines directions for future studies. Supporting information, including model assumptions and diagnostic checks, is provided in Section 6: Appendix.

2. Data

2.1 Data source

The study draws on The Washington Post School Shootings Database (citation), a systematically compiled record of firearm incidents occurring at U.S. primary and secondary schools during school hours since the Columbine High School shooting on April 20, 1999. The database was developed by integrating multiple information channels, including Nexis archives, news reports, open-source databases, law enforcement records, school websites, and direct communications with school and police officials. Incidents were included only if gunfire occurred on campus immediately before, during, or after classes, thereby excluding after-hours events, accidental discharges injuring only the shooter, and suicides that posed no threat to others. Since no federal agency tracks school shootings, The Washington Post database represents one of the most comprehensive and systematically verified datasets available for studying these events. It further incorporates school-level demographic and enrollment information from the U.S. Department of Education's Common Core of Data and Private School Universe Survey, allowing for contextual analyses of how school characteristics relate to the severity of shooting incidents.

2.2 Data overview

The raw data consists of 428 recorded school shooting incidents in the United States, spanning the period from April 20, 1999 onward. Each observation represents a distinct shooting incident, and the dataset contains 50 variables. These variables include temporal and geographical information about the shooting incident, shooter(s) demographic characteristics, shooter(s) relationship to school community, shooting type, shooting outcomes, weapon used and source, school type, number of students at school(total enrollment and by racial groups), number of students at school eligible to receive a free or reduced-price lunch, presence of a school resource office, and grade-levels offered by school. The dataset records a maximum of two shooters per incident. Every incident has at least one shooter, and details for a second shooter are included only if applicable. The full list of variables can be found in the Appendix.

2.3 Data cleaning

To address inconsistencies and typos in categorical variables, we cleaned and standardized their string representations. For “shooting type”, the original categories included “unclear”, “public suicide”, “hostage suicide”, “accidental” and “targeted.” To simplify analysis and ensure sufficient sample sizes within each group, we combined “public suicide” and “hostage suicide” into a single “suicide” category, as these types were rare (8 and 2 observations, respectively). One observation labeled “public suicide (attempted)” was also included under “suicide.” Six observations labeled “targeted and indiscriminate” were grouped under “targeted,” and two labeled “accidental or targeted” were grouped under “accidental”. Missing values were assigned as “unknown”. We also re-coded “shooter’s relationship to the school community” into broader, interpretable categories. Specifically, we grouped the textual entries into six main categories: “Current Student”, “Former Student”, “Family/Guardian of Student”, “Non-Security Staff”, “Family of Staff”, and “Police/Security”. For example, descriptions such as “student in summer school” and “suspended student” were recoded as “Current Student” and “Former Student,” respectively. All remaining or missing entries were categorized as “Other, Unknown or No Connection”. This standardization resolves inconsistent textual entries, and facilitates simpler comparisons across levels embedded in the categorical variables in subsequent analyses.

One observation had shooter 1’s gender recorded as “h,” which was presumed to be a data entry error. Since this case did not involve a second shooter and the intended gender could not be inferred from other available information, the observation was excluded from the dataset. Three observations had number of students eligible for free or reduced-price lunch greater than total enrollment, they were removed.

Below is a summary of (non-zero) count and proportion of missing values in our variables of interested after basic cleaning.

Table 1: Count and Proportion of Missing Values in Interested Variables

Variable	Missing Count	Missing Proportion (%)
Age of shooter 1	127	29.95
Gender of shooter 1	94	22.17
Age of shooter 2	412	97.17
Gender of shooter 2	412	97.17
Number of white students	10	2.36
Number of students eligible for free/reduced-price lunch	56	13.21
Resource officer present	187	44.10
Weapon used	170	40.09
Weapon source	321	75.71
Shooter 2 relationship with school community	418	98.58

According to Table 1, more than 97% of incidents did not involve a second shooter, indicating that the vast majority of incidents are perpetrated by a single individual. Variables “presence of a resource officer”, “weapon used”, and “source of the weapon” contains missing values for over 40% of the observations. The high proportion of missing values, combined with the specificity uncoded in weapon-related variables, makes reliable imputation impractical. Therefore, in the following analysis, we exclude all information related to a second shooter, as well as presence of a resource officer and weapons.

To handle missing values, we applied context-specific imputation strategies for the remaining key variables. For shooter 1’s age, missing values were imputed by the mean age within the same shooting type and state, with state- or overall means used as fallbacks. Shooter 1’s gender was imputed probabilistically based on the observed state-level gender proportions, preserving underlying shooter gender distribution in the data. No fallback is needed here since male shooter proportions are available in all states. For the number of students eligible for free or reduced-price lunch, we first set all private schools to zero, as such programs typically apply only to public schools. For public schools, missing values were imputed using the average lunch proportion within each state, with the overall proportion used as a fallback. The imputed value was then scaled by the school’s total enrollment. Finally, missing counts of white students were imputed using the mean proportion of white students within each combination of shooting type, state, and school type and multiplied by school’s total enrollment. Mean at state–school type level, and school type mean were used as fallbacks.

We collapsed the number of killings and injuries variables into binary indicators. For observations with number of killed or injured greater than zero, the corresponding indicator was coded as “1”; otherwise, it was coded as “0”. This is appropriate because our study focuses on the presence or absence of harm in each incident, rather than the exact number. This approach hence simplifies the analysis, and facilitates clear interpretation of the relationships between incident characteristics and the occurrence of harm. Additionally, we created two standardized variables to capture key aspects of students’ socioeconomic and racial composition in each school. The racial composition was measured as the proportion of non-white students calculated as $1 - \frac{\text{Number of white students}}{\text{Total enrollment}}$, and the economic situation was measured as the proportion of students receiving subsidized meals calculated as $\frac{\text{Number of students eligible for free or reduced lunch}}{\text{Total enrollment}}$. Standardizing both measures by total enrollment allows for meaningful comparisons across schools of different sizes and ensures that these variables reflect relative composition rather than absolute counts.

R version and packages (citation)

2.4 Data description

The cleaned data consists of 424 observations and 9 variables. Below is a summary description of all variables in the cleaned dataset.

Table 2: Summary Description of Variables

Variable	Type	Categories/Range
Killed	Binary	1 = At least one killing (excluding shooter) occurred
Injured	Binary	1 = At least one injured (excluding shooter) occurred
School Type	Categorical	Public, Private
Shooting Type	Categorical	Indiscriminate, Targeted, Accidental, Suicide, Unclear
Shooter Age	Numerical	6 to 74
Shooter Gender	Categorical	Male, Female
Shooter Relationship	Categorical	Current Student, Former Student, Family/Guardian of Student, Non-Security Staff, Police/Security, Other/Unknown/No Connection
Proportion of Non-white Students	Numerical	0 to 1
Proportion of Students Eligible for Subsidized Lunch	Numerical	0 to 1

2.5 Data summary and visualizations

In total, the dataset includes 424 shooting incidents. Among them, 328 cases involved no killings, while 99 cases resulted in at least one fatality. This distribution indicates that the majority of school shootings (approximately 77%) did not involve any deaths.

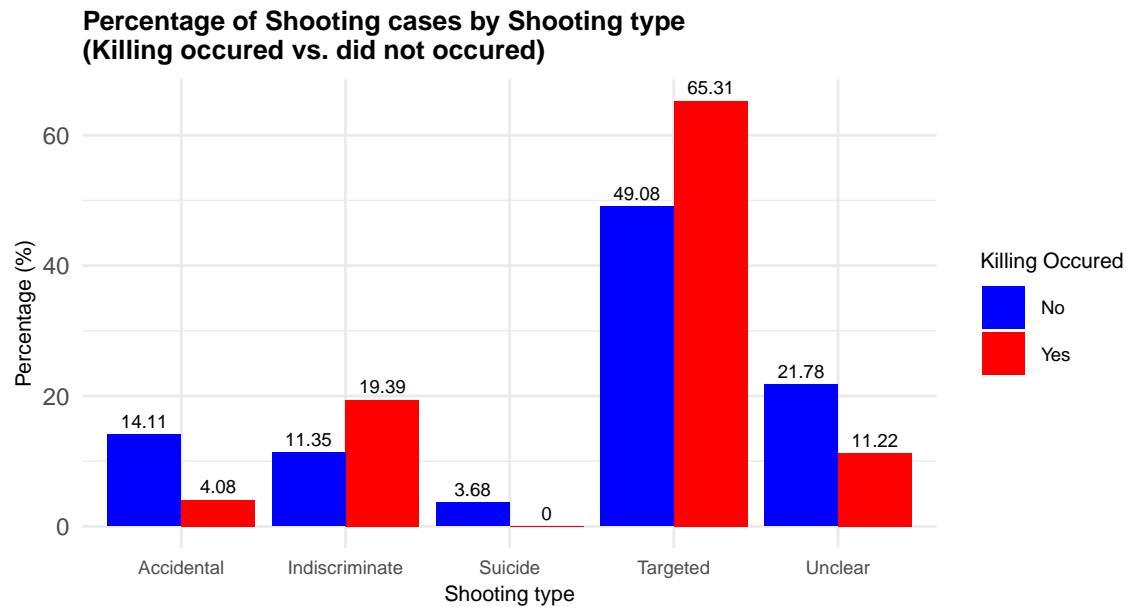
Table 3: Shootings with Killings

Measure	Shooter Age	Proportion of Non-white Students	Proportion of Students with Subsidized Lunch
Mean	20.41	0.57	0.48
Standard Deviation	9.63	0.35	0.30
Median	17.00	0.60	0.48
Min	6.00	0.00	0.00
Max	53.00	1.00	1.00

Table 4: Shootings without Killings

Measure	Shooter Age	Proportion of Non-white Students	Proportion of Students with Subsidized Lunch
Mean	19.48	0.69	0.58
Standard Deviation	8.74	0.33	0.29
Median	17.00	0.88	0.60
Min	6.00	0.00	0.00
Max	74.00	1.00	1.00

Tables 3 and 4 reveal that shootings involving killings were associated with a slightly older average shooter age. In contrast, non-fatal shootings occurred in schools with higher proportions of non-white students and students receiving subsidized lunch, suggesting that non-fatal incidents were more common in schools with socioeconomically disadvantaged or racially diverse students. However, the standard deviation were substantial for all variables, suggesting considerable heterogeneity even within the incident groups. Please refer to Appendix for additional visualizations.



3. Model

5. Conclusion

These incidents may be less predictable because they occur outside existing social or disciplinary networks — schools can't rely on familiar early-warning signs like student behavioral issues, conflict reports, or peer concerns. Such shooters may enter the school with higher intent or preparation, possibly bringing more lethal weapons or planning attacks more deliberately. Because they are external actors, existing school-based prevention systems (focused on students and staff) may be less effective in identifying or deterring them beforehand. Schools can't easily prevent these attacks directly — but it highlights a different area of vulnerability and thus a different kind of prevention need: It suggests that school safety cannot rely solely on internal behavioral monitoring (e.g., identifying at-risk students). There is a need for stronger external security coordination — for instance, improving access control, visitor screening, and information sharing between schools, law enforcement, and community organizations. It also underscores the importance of rapid emergency response and preparedness training, since these events tend to be sudden and high-lethality when they occur.

Appendix

Raw data variables

- Unique identifier for each shooting incident

- National Center for Education Statistics (NCES) unique school ID
- Name of the school
- NCES unique district ID
- Name of the school district
- Date of the shooting
- School year of shooting
- Year of the shooting
- Approximate time of the shooting
- Day of the week when the shooting occurred
- City where the school is located
- State where the school is located
- Type of school (Public or Private)
- Total enrollment at the school at the time of the shooting
- Number of people killed (excluding the shooter)
- Number of people injured (excluding the shooter)
- Total number killed and injured (excluding the shooter)
- Type of shooting: accidental, hostage suicide, indiscriminate, public suicide, targeted, or unclear
- Age of the first shooter
- Gender of the first shooter (Male or Female)
- Race or ethnicity of the first shooter (American Indian, Black, Hispanic, White, Asian)
- First shooter's relationship to the school
- Indicates if the first shooter died in the incident (yes or no/unknown)
- How the first shooter died (if deceased)
- Age of the second shooter (if any)
- Gender of the second shooter
- Race or ethnicity of the second shooter
- Second shooter's relationship to the school
- Indicates if the second shooter died (yes or no/unknown)
- How the second shooter died (if deceased)
- Number of White students enrolled at the time of shooting
- Number of Black students enrolled
- Number of Hispanic students enrolled
- Number of Asian students enrolled
- Number of American Indian or Alaskan Native students enrolled
- Number of Hawaiian Native or Pacific Islander students enrolled (unavailable before 2009)
- Number of students of two or more races (unavailable before 2009)
- Presence of a school resource officer or security guard (present or absent/unknown)
- Weapon(s) used in the shooting
- Source from which the shooter acquired the weapon(s)
- Latitude of the school
- Longitude of the school
- Full-time equivalent teachers at the school at the time of shooting
- Lowest grade level offered by the school
- Highest grade level offered
- Number of students eligible for free or reduced-price lunch
- County where the school is located
- Two-digit state Federal Information Processing Standards (FIPS) code
- Five-digit county FIPS code
- NCES urban-centric locale code