Time Series Analysis of School Shootings in the United States and Forecast for the 2019-2020 School Year

David J Corliss

Peace-Work, Plymouth, MI

ABSTRACT

This study presents a historical analysis of school shooting attacks in the United States. Only attacks using firearms were included: raw data from the US federal CHDS school shooting database was analyzed to exclude suicides, accidents, and threatening actions not resulting in injury. Exploratory data analysis indicated two historical time periods with distinct behaviors. Time series cluster analysis indicated a change in attack behavior about the year 1991: prior to that time, attacks and fatalities were relative few, normally distributed, and show little variation from year to year. Beginning about 1991, a second type of attack pattern is found superimposed on the historical pattern, with both continuing to the present. This second pattern is found to be consistent with stochastic terrorism, characterized by a highly variable annual attack rate, marked increases in deaths and lethality (deaths per incident), and a skewed distribution with a risk of a high number of fatalities driven by a small number of extreme events. A school shooting forecast for the 2019-2020 school year is estimated with 17±7 incidents and 15±10 fatalities. This analysis was performed using SAS University Edition.

INTRODUCTION: SCHOOL SHOOTINGS IN THE UNITED STATES

While there has been increasing concern over gun violence over the years, especially in the area of school shootings, research in this area is very incomplete. Tragic events, including mass shootings at schools in Columbine in Aurora, CO, Sandy Hook, CT, and Stoneman Douglas HS in Florida, have not translated into a substantial body of rigorous scientific research. Historically, questions related to general threats to the physical well-being of the public at large have been investigated by publicly-funded research organizations such the Centers for Disease Control and Prevention in Atlanta (CDC, https://www.cdc.gov/). However, there is substantial resistance in some circles to prevent research into gun violence in the US. In 1996, an amendment in the US federal government omnibus spending bill established the requirement that "none of the funds made available for injury prevention and control at the Centers for Disease Control and Prevention (CDC) may be used to advocate or promote gun control" (see Public Law 104–208, 104th Congress). At that time, the amount of funding previously budgeted to the CDC for gun violence research was re-directed toward other projects.

The intent and exact legal requirements of this regulation, known as the Dickey Amendment after its author, US Representative from Arkansas Jay Woodson Dickey, Jr., have been debated. However, the practical effect of the Dickey Amendment has been to prevent any federally-funded research in the area of gun violence. While this has left a significant void in the area of gun violence research, NGO's may be unaffected by government restrictions. This study was produced by the NGO Peace-Work, an all-volunteer cooperative of statisticians, data scientists and other researchers applying statistical modeling and methods to issue-driven advocacy, including poverty, education, human trafficking, and social justice.

SAS University Edition as a powerful analytic tool with extensive statistical, machine learning, and big data capabilities that is free for non–commercial use. Using virtual server technology, it provides a familiar console-based SAS coding environment with an extensive set of SAS procedures for data management, analysis, and visualization. Peace-Work expresses its deep appreciation to SAS for making University Edition available for this research.

DATA DEVELOPMENT

Source data for this study comes from the K-12 School Shooting Database, a project of the United States Center for Homeland Security Defense and Security with the Federal Emergency Management Agency (FEMA), This compilation, undertaken at the Naval Graduate School in Monterey, California, captures information on K-12 school shootings in the United States from 1970 forward. Information is taken from newspaper and magazine articles, academic journal reports and case studies, law enforcement, internet searches, and other sources. The project seeks to collect a wide range of data on school shootings, including date, location, victim numbers, information about the school, whether suicide occurred or was attempted, type of incident (e.g., racially motivated, accidental etc.) and other fields. Each incident is assigned a reliability rating on a scale of 1 (lowest) to 5 (considered most reliable).

This publicly available database is continuously updated, containing 1,405 records as of September 1st, 2019. Each record in the database comprises a single incident. The database contains no other resources, performs no analysis, and makes no conclusions or recommendation: it is exclusively a compilation of data identified in searches. This study by Peace-Work seeks to analyze attacks, defined as deliberate acts where persons other than the shooter were wounded or killed, whether the shooter was harmed or not.

Data privacy was maintained at all times. While these data are taken from a publicly available source, no personally identifying information in the CHDs database was read into a SAS dataset for this analysis.

Data were aggregated by school year prior to analysis. School year was extracted from the date on each record, with a year defined as running from August 1st through July 31st. To identify deliberate attacks resulting in victims with physical injuries, five selection criteria were applied to the raw data records in the Schools Shooting Database:

- As the analysis was performed at the School Year level, partial school years were excluded from this analysis – specifically, records prior to 8/1/1970 and after 7/31/2019.
- Suicides where the shooter was the only victim were excluded. These records are identified in the first of four fields regarding suicide in the CHDS database.
- Accidental shootings were excluded: Category = 'Accidental' in the CHDS data
- Unreliable incident reports were excluded: Reliability = 1 in the CHDA data, described as "Independent Single Author/Moderator Blog, report/list lacking citations, or cited source cannot be located".
- Threats where no person was physically harmed were excluded: killed + wounded = 0 => drop the record.

ANALYSIS

Exploratory data analysis indicated two different historical periods. Early years in the time series show relatively fewer incidents, fatalities, and variability in these characteristics. This is called the Historical pattern. Beginning in the early 1990's, a sharp increase is seen in school shooting incidents and fatalities. These new records in the Modern era possess distinct characteristics, including more deaths per incident, greater variability from year to year, and a highly skewed distribution with a long tail where the historical years presented a normal distribution. The resulting pattern, since the early 1990's, is a combination both Historical and Modern-type attacks (Figure 1).
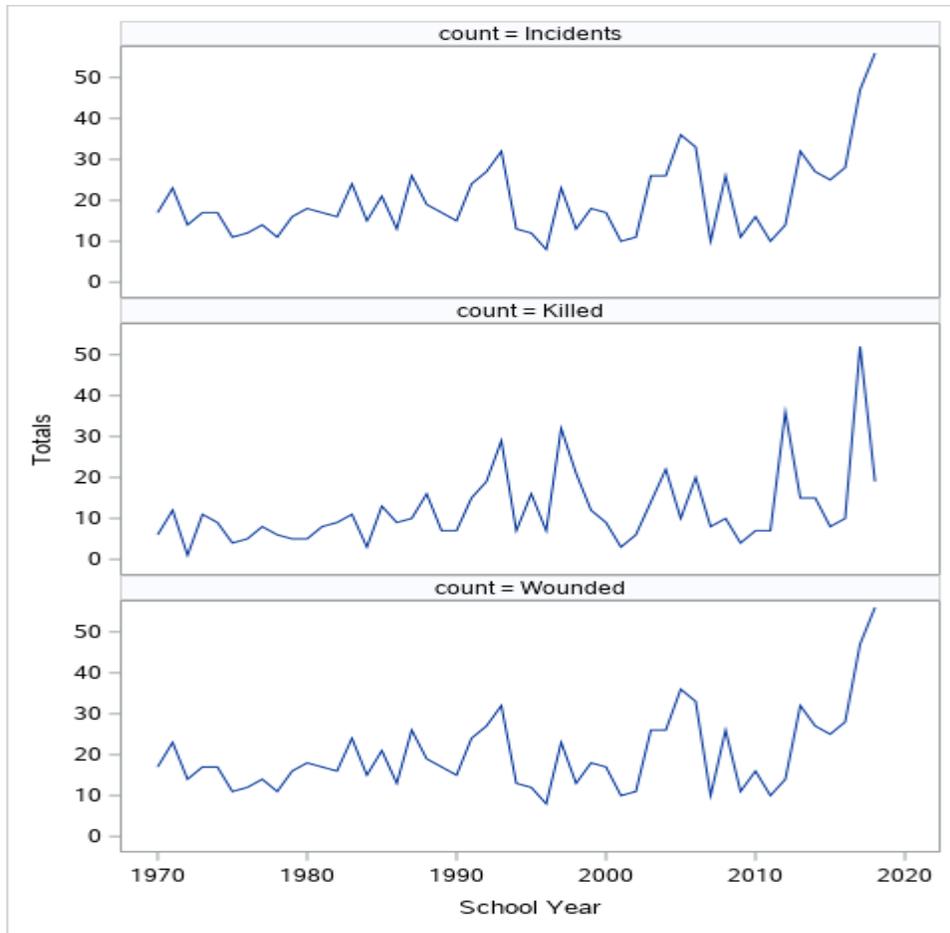
Figure 1: Historical Patterns in US School Shootings by School Year

Time series cluster analysis using PROC FASTCLUS was used to identify distinct periods in the history. Following best practices for this method, both totals at each point in time and rate of change variables were considered in the analysis. Incidents per school year, number killed, and absolute magnitude of the changes for each gave the clearest separation of the clusters:

```
proc fastclus data=pw.gv_annual_totals maxc=3 maxiter=10 out=work.cluster1;
    var school_year incidents incidents_abschange killed killed_abschange;
run;
```

The analytic design calls for initially creating more clusters than needed and then pruning the result. Three time series clusters were identified: (1) 1970-71 to 1990-91, (2) 1991-92 to 2016-17, and (3) 2017-18 to 2018-19. With just two years as the very end placed in a separate cluster – effectively, an outlier – the third cluster was combined with the second. Comparison of the summary statistics prior to the 1991-1992 school year to the following period – a superposition of Historical and Modern time series allows calculation of these statistics for the Modern series only. Properties of the two type are summarized in Table 1.

Historic Properties: Through the 1990-1991 School Year

Variable	N	Median	Mean	Std Dev
Incidents	21	17.00	16.81	4.04
Killed	21	8.00	7.86	3.58
Wounded	21	18.00	26.48	20.89
Killed per Incident	21	0.47	0.47	0.17
Wounded per Incident	21	1.06	1.58	0.96

Modern Plus Historic Properties: The 1991-1992 School Year Through the Present

Variable	N	Median	Mean	Std Dev
Incidents	28	23.50	22.54	11.72
Killed	28	13.00	15.46	10.98
Wounded	28	30.50	30.21	23.30
Killed per Incident	28	0.55	0.69	0.50
Wounded per Incident	28	1.30	1.34	0.57

Calculation of Modern Pattern Only: The 1991-1992 School Year Through the Present

Variable	N	Median	Mean	Std Dev
Incidents	28	6.5	5.73	12.53
Killed	28	5	7.6	11.96
Wounded	28	12.5	3.73	12.06
Killed per Incident	28	0.77	1.33	0.54
Wounded per Incident	28	1.92	0.65	-0.15

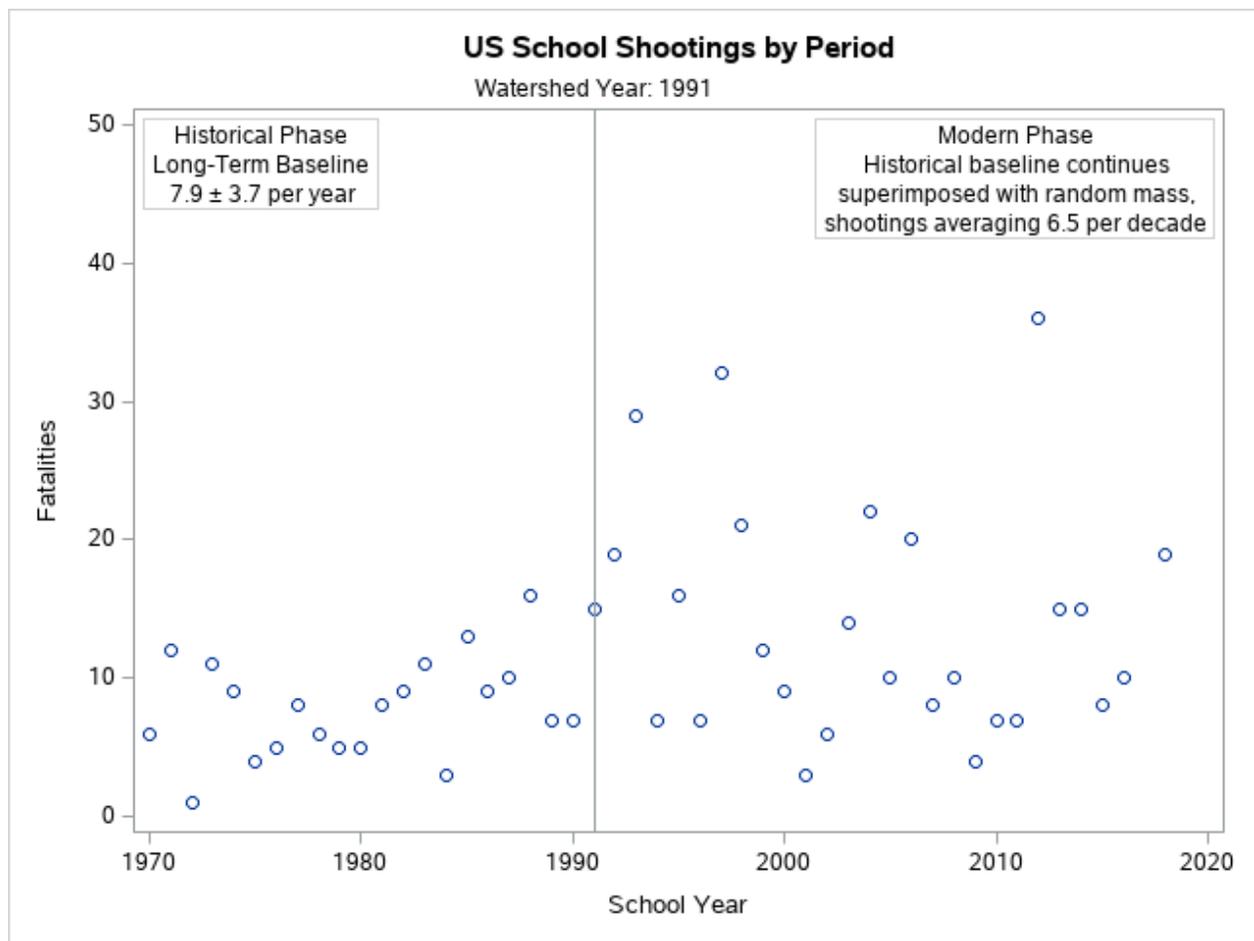Table 1: Summary statistics on US school shootings from successive time series clusters.

Figure 2: US School Shootings by Historical Period. Records in the source data excluded here are accidental shootings, suicides where only the shooter was harmed, incidents where no one was killed or wounded, and data from uncorroborated reports. Source Data: CHDS. Analysis: Peace-work.

Here is the source code for the plot:

```
proc sgplot data=pw.gv_annual_totals;
    scatter x=school_year y=killed;
    xaxis values=(1970 1980 1990 2000 2010 2020) label='School Year';
    yaxis values=(0 10 20 30 40 50) label='Fatalities';
    refline 1991 / axis=x label='Watershed Year: 1991';
    inset 'Historical Phase' 'Long-Term Baseline' '7.9 ± 3.7 per year' /
          position = topleft valuealign=center border;
    inset 'Modern Phase' 'Historical baseline continues'
          'superimposed with random mass,' 'shootings averaging 6.5 per decade',
          position = topright valuealign=center border;
    title 'US School Shootings by Period';
run;
```

STOCHASTIC TERRORISM

- Define stochastic terrorism

5

- Discuss the data: the time series of school shootings matches the growth of internet chat rooms and, later, social media and their documented ability through "echo chambers" to reinforce extreme views and foster action
- Role of internet echo chambers in stochastic terrorism

The time series cluster analysis used in this study initially found three clusters; the third cluster consisting of the last two complete school years in the data was combined with the second, similar interval beginning in 1991-1992. Alternatively, if the stochastic terrorism hypothesis is correct, it should be pointed out that increase seen in the last two years In the data are consistent with a documented (Corliss, 2018) rise in hate speech sources in Twitter during this same period (Figure 3).
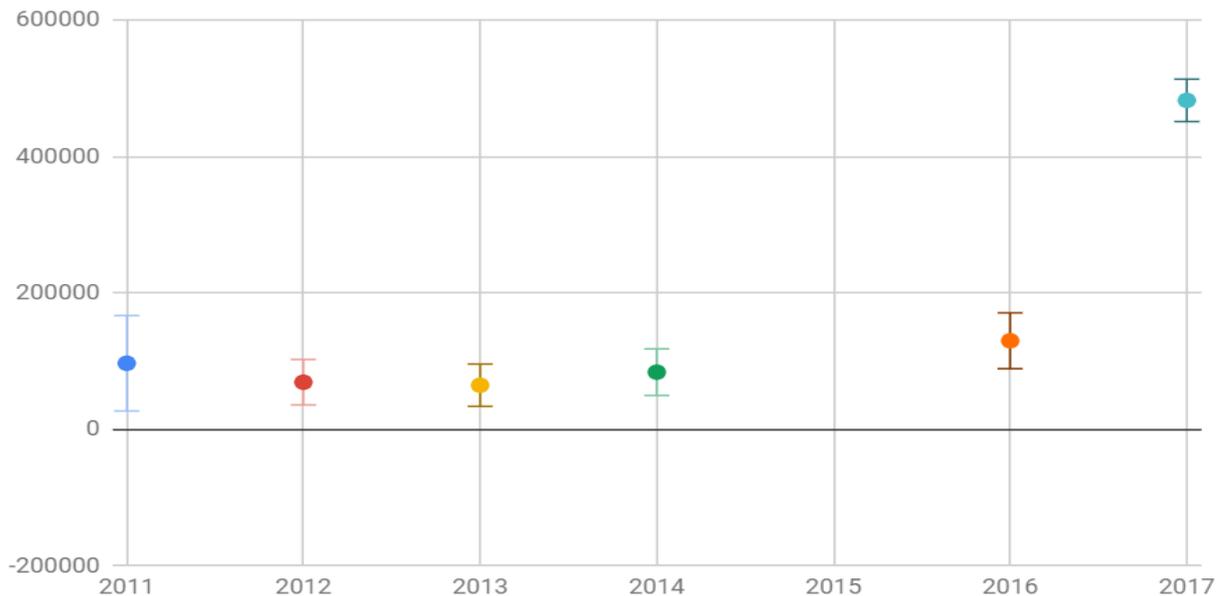

Figure 3: Time Series of Hate Speech Source in Twitter, 2012-2017 (Corliss, CIRM, 2018)

One noteworthy property of these data is something _not_ observed: there is no apparent effect of the US federal assault weapons ban. Neither a decrease when instituted in September 1994 nor an increase following its expiration 10 years later is observed. Several factors may have affected this outcome, including but not limited to

- The ban was on *manufacture*, not ownership or use
- Wide availability of weapons included under the ban
- Limited scope of the ban – many highly lethal and / or high rate of fire weapons were not included. For example, the perpetrators of the 1999 Columbine High School mass shooting used several different weapons, none of which were included in the assault weapons ban.

CONCLUSIONS

Time series cluster analysis of school shooting incidents in the CHDS school shooting database identifies two historical periods. This analysis only included deliberate attacks, excluding accidents, suicides without other victims, and unreliably reported events. In the first time period, from the beginning of the data in 1970 through about the 1990-1991, there was an average of 16.8 ± 4.0 incidents and 7.9 ± 3.6 deaths per school year. Data per year in this period are normally distributed.

Beginning in the early 1990's, a second type of attack was observed in addition to a continuation of the historical pattern. These modern attacks, continuing to the present, have an average of 5.7 incidents and 7.6 deaths per school year. Variability increased substantially in a highly skewed distribution, with the standard deviations exceeding the means at 12.5 for incidents and 12.0 for deaths. The Modern pattern of attacks is far more lethal (average 1.33 killed per incident) than the Historic (0.47 killed per incident). The mean number killed per incident increased more than 280% from the Historic period, even as the number wounded fell by half. Analysis of the time series does not support the finding of any effect from the US ban on the assault weapons manufactured from September 1994 through September 2004. However, the timing of the rise of the modern-type attacks is consistent with the growth of internet, where a number of perpetrators were very active. As often described by perpetrators and consistent with the observed time series, breeding ground for the modern type of attacks is found first in internet chat groups and later through social media. This hypothesis for the sea change in school shooting observed in recent decades involves perpetrators being motivated and encouraged personal but virtual connections – a phenomenon that could be termed *micro-hate* - and further incited by the *macro-hate* of stochastic terrorism. Investigation of this hypothesis is recommended.

The forecast for the 2019-2020 school year 11 to 24 school shooting incidents, with a slight (7%) chance of a peak year with more than 30 incidents. Most of these will result in two or fewer fatalities and will not be picked by national media reporting services. 5 to 21 deaths are forecast, with a slight (7%) chance of a peak year with more than 35 deaths. In the event of a peak year for deaths, it will likely be driven by one or two very large-scale mass shootings.

ACKNOWLEDGEMENTS

REFERENCES

(Reference needed for stochastic terrorism)

Corliss, D. J. (2012), "Dynamically Evolving Systems: Cluster Analysis Using Time", Proc. SAS Global Forum 2012

Corliss, D. J. (2018), "Bayesian Capture Recapture in Social Justice Research", CIRM

Riedman, D. and O'Neill, D. (2018), K-12 School Shooting Database, Naval Postgraduate School Center for Homeland Defense and Security, Retrieved 8/11/2019 from https://www.chds.us/ssdb/

Tryon, R. C., 1939, Cluster analysis. Ann Arbor: Edwards Brothers

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

David J Corliss Peace-Work
davidjcorliss@peace-work.org