

SOC 504 Reproducibility Memo

David Liu and Sri Nimmagadda

March 29, 2018

Background

We are replicating the results of a paper titled *Firearms and accidental deaths: Evidence from the aftermath of the Sandy Hook shooting* [1]. The paper appeared in the December 2017 issue of Science and was authored by two economists from Wellesley College. We chose this paper to replicate because it is both recent and socially relevant as debates over gun control captivate the country.

The paper's main contribution was establishing a casual link between "gun exposure" and accidental deaths. In the article, gun exposure is measured both in terms of online activity, such as google searches relating to firearms, and in terms of gun purchases. An example of an "accidental death" would be an unintentional fatal firing from a firearm. To establish the causal claim, the authors first collected health and firearm data from a variety of government entities. Then, Levine and McKnight utilized fixed effects and instrumental variables in their casual inference analysis. The paper's statistical content relating to causal treatment effect overlap heavily with this course and SOC 500.

This memo will enumerate the specific replication objectives, recap our data collection process, and compare our results against those published in the paper. Finally, we propose extension ideas for feedback.

Replication Targets

It may be helpful to explicitly specify the values and quantities that we aim to replicate. All of the paper's claims are captured in its three figures and single table. The figures contain contextual data that setup the causal inference analysis, while the table contains the regression outputs, the crux of the paper. Below, we will enumerate these four entities and their significance:

1. **Figure 1: Google Trends.** Using data from Google Trends, the authors argued that immediately after the Sandy Hook shooting, people had a greater interest in either purchasing a firearm or restoring an old one. For our replication, we will re-generate the two trend lines and compare them with the figure.
2. **Figure 2: Association between exposure and death** The most important figure in the paper, the second figure establishes a correlation between exposure and accidental deaths. On the same plot, the authors overlay trends in gun purchases as well as accidental deaths. It will be important to confirm that exposure and accidental deaths both spike following Sandy Hook.

3. **Figure 3: State-level trends in exposure** For visual clarity, the authors provided a geographical map highlighting states with high spikes in gun purchases following Sandy Hook. We will reproduce this state-level map.
4. **Table: Causal inference analysis** This is the most important replication target. The authors perform causal inference for two subpopulations (children and adults) and for each subpopulation, use two separate methods (OLD and IV) to establish casual effect. We will verify the values of these regression coefficients.

Restricted data

This paper did utilize restricted data which we were not able to obtain. Nevertheless, we were able to replicate the paper's primary focus because the restricted data only supplemented the main findings. Specifically, the authors demonstrated that, at a national level, the Sandy Hook shooting was followed by greater gun exposure, which caused additional accidental deaths. They then performed the same analysis at the state level, to further back up their causal claim. However, state-level mortality data is restricted so we were not able to replicate Figure 4 and Table 1, Panel 4 from the paper.

Data Collection

The original paper utilized four datasets for its analysis. These datasets originated from Google Trends, the National Instant Criminal Background Check System (NICS), the National Cancer Institute's Surveillance, Epidemiology, and End Results program, and the CDC. Fortunately for us, the first three of these datasets were posted on the article's Dataverse. Then, following directions in the Dataverse README, we were able to obtain the mortality data from the CDC. The mortality data was most difficult to manage because the files were 10 GB in total. Otherwise, Dataverse enabled a relatively seamless data collection process.

Comparison of Results

Figure 1: Google Trends

The search trends figure was the most straight forward to replicate because it only utilized one data set. The Dataverse did not contain code to replicate the figure directly, but we were able to import the data into R and generate the figure with ggplot. A side-by-side comparison of the two graphs are shown in Figures ?? and 2.

The weeks on our x-axis are still quantified in an obscure manner, the but overall trends, especially the spike following Sandy Hook, match. The original article documentation states that Google Trends data itself may be unstable depending on the time of query. We have not yet assessed this claim but plan to do so.

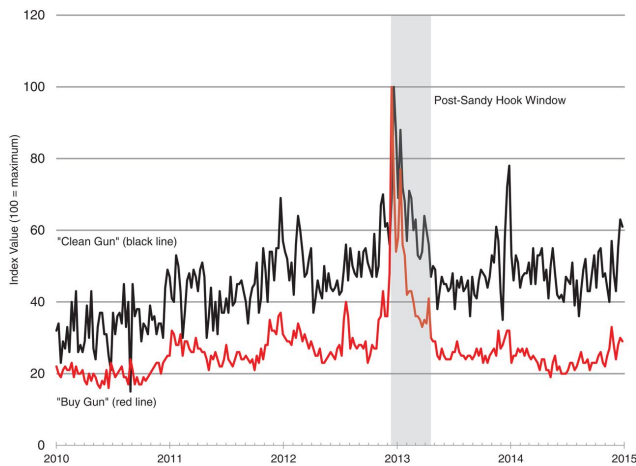


Figure 1: Original graph

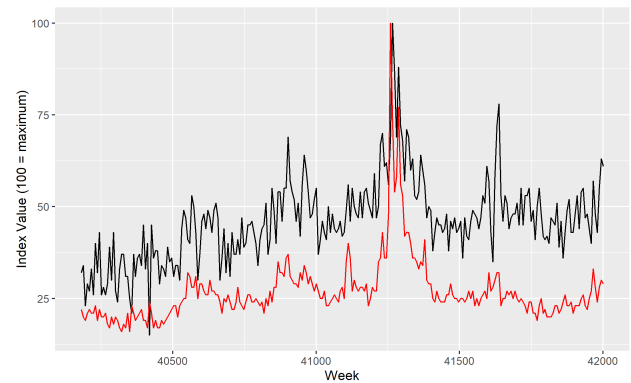


Figure 2: Generated graph

Figure 2: Association between exposure and deaths

This figure establishes a correlation between firearm exposure and accidental deaths, setting the stage for the causal inference that occurs later. In the replication process, we encountered a few discrepancies with the reported figures and code. We will discuss these differences in depth.

For context, the original figure label states that the data are "deviations" from the expected values. Taking a look at the code, this wording translates into fitting regressions for firearm sales and accidental deaths, individually, and then plotting the residuals. For the firearm sales the authors fitted a fixed effects regression that accounts for monthly seasonality and yearly trends. And for the accidental deaths data, the authors fitted a quadratic function of deaths on time. There does not appear to be any documentation on why a quadratic trend was shown.

After running the author's code, we developed the plot shown in Figures 3 and 4. For reference, the original plot is shown in Figure 5.

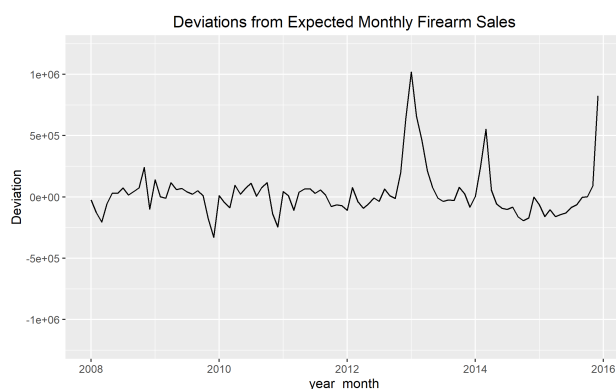


Figure 3: Sales deviations

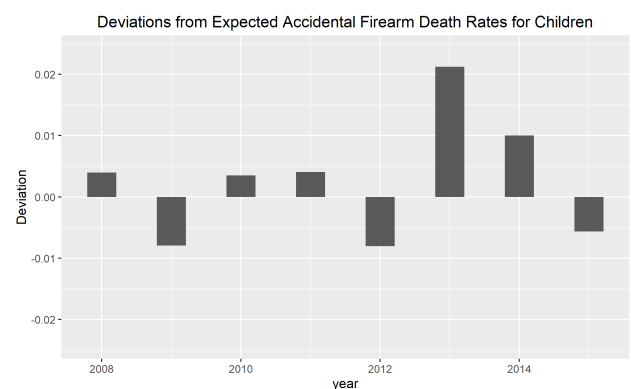


Figure 4: Firearms deaths deviations

Discrepancies with author's results

In order to create Figures 3 and 4, we discovered a minor bug in the author's Stata code: when building the fixed-effects regression, they mistakenly included the 2016 firearms data. In fact,

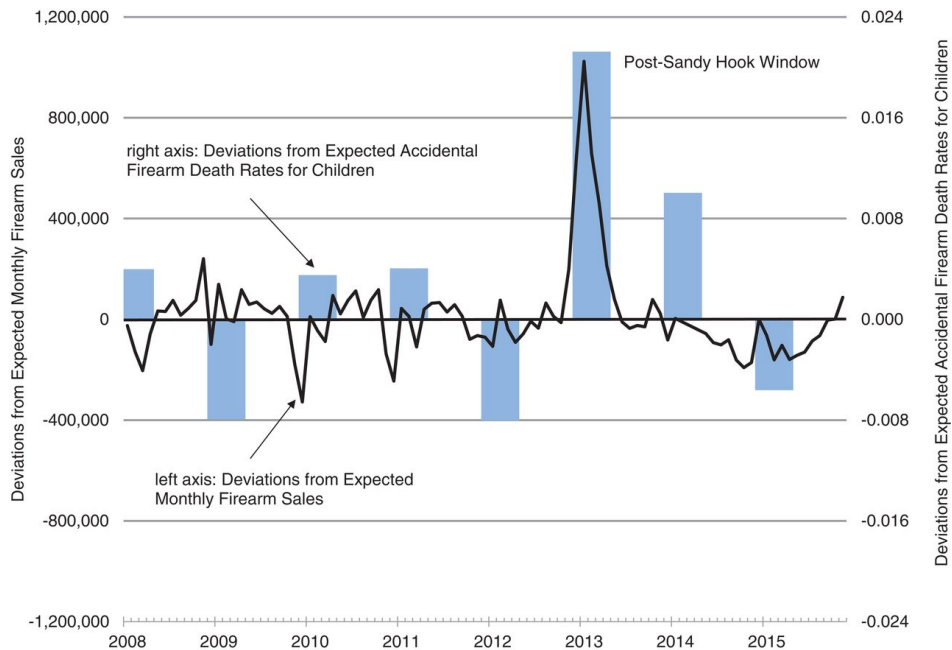


Figure 5: The original figure from the paper.

when running the Stata code as-is, we generate the plot shown in Figure 6. It is evident that including the 2016 data distorts the estimates for 2008 because both years must share a single fixed effects coefficient.

Even after excluding the 2016 data, we see that the firearms deviation line plot differs from the

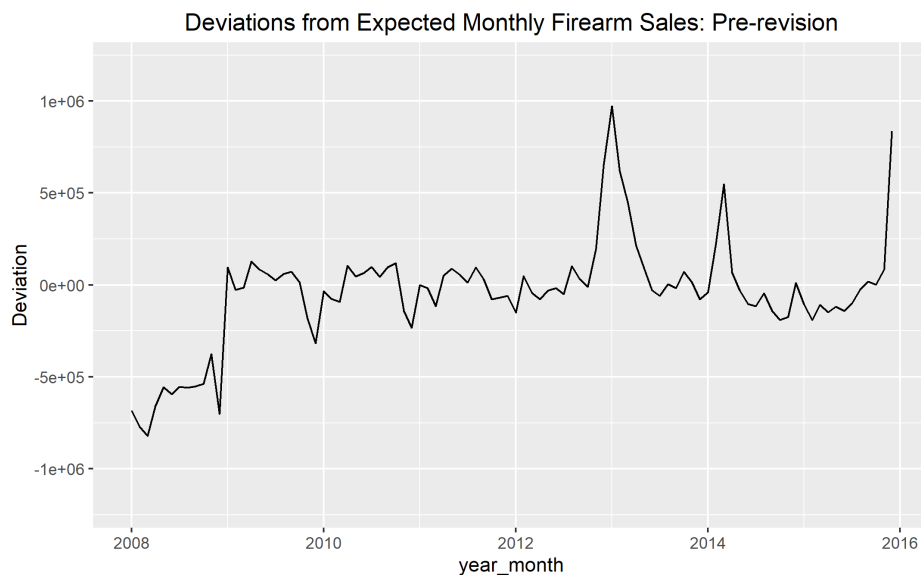


Figure 6: The original Stata code regressed on 2016 data which distorted estimates for 2008, the intended omitted variable.

one in the paper. Specifically, there is a noticeable peak in 2014 and the final data point, December 2015 is much higher in our plot. These values undermine the author's claim that the peak in firearm sales was unique to the Sandy-Hook area. We are still investigating the source of the discrepancy.

Figure 3: State-level trends in exposure

As a precursor to their state-level analysis, the authors demonstrated the heterogeneity in gun sale increases by state. To calculate these numbers, the authors first normalized gun purchases by state population. Then, for each state, the authors fit a fixed effects regression of gun sales on a dummy variable, which is one in the months following the shooting and zero otherwise. As before, the regression contained dummy variables to account for seasonality and annual trends. They then defined the increase in gun sales to be the coefficient on the Sandy-Hook dummy variable.

To replicate, we simply ran the provided Stata code, which outputted a slew of regression coefficients. We then manually transferred these coefficients into a CSV data file, which maps states to gun sale increases. We then used ggplot's `geom_map` geometry to visualize. The side-by-side comparison is shown in Figures 7 and 8. Though a few tweaks still need to be made, overall, these two maps match precisely.

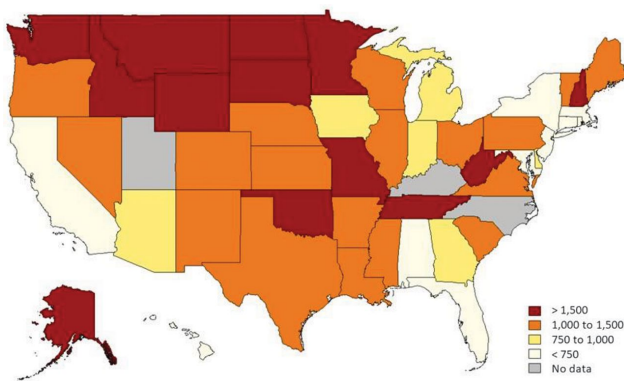


Figure 7: Original map

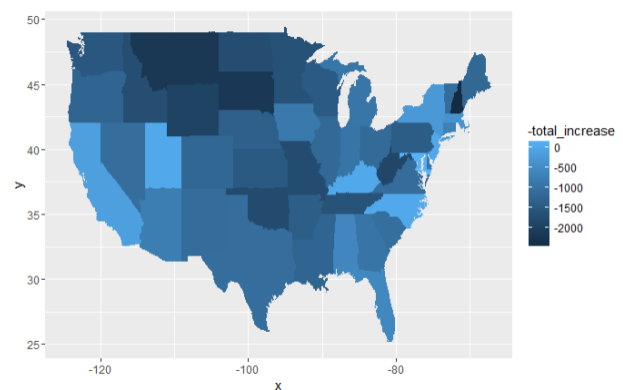


Figure 8: Generated Map

Causal inference analysis

Interestingly, the paper's most significant findings were the most straight forward to replicate as many of the quantities were simply regression coefficients. To start, the original quantities are shown in Figure 9.

These first panel of the table provides descriptive statistics, such as total population and average number of accidental fire arm deaths. We ran the provided Stata code to yield the values shown in Figure 10. The mean values of `pop_byage` match the population parameters from the table. Then, if we multiply the average value of `num_deaths` by five, to obtain the five-month mortality rate, we also get the average values reported in the paper. To make causal conclusions, the authors first ran an OLS regression with a dummy variable for the Sandy Hook period representing a treatment effect. To assert that exposure produced more accidental deaths, the authors then performed an IV regression with 2 stage least squares, where the Sandy Hook shooting was the instrument and the amount of exposure was the endogenous variable.

In all six of these regression (two methods each applied to three age groups), we must verify the coefficient on `sandyhook`. To be concise, we do not copy those regression outputs here, but they are contained in the `replication_science.log` log file. The coefficients in the final six regressions do match those reported in the paper.

	All ages	Children (ages 0 to 14)	Adults (ages 15+)
<i>Descriptive statistics</i>			
Average number of accidental firearm deaths in a 5-month window	210	28	182
2013 population (in millions)	316	61	255
<i>National data, 2008–2015, ordinary least squares</i>			
Impact of Sandy Hook	0.0036*** (0.0013)	0.0058*** (0.0021)	0.0031* (0.0016)
Implied number of additional accidental firearm deaths	57	18	40
<i>National data, 2008–2015, instrumental variables</i>			
Impact of 1000 gun sales per 100,000 population	0.019*** (0.006)	0.030*** (0.010)	0.016** (0.007)
Implied number of additional accidental firearm deaths	57	17	39
<i>State data, 2008–2015, instrumental variables</i>			
Impact of 1000 gun sales per 100,000 population	0.022** (0.010)	0.038** (0.018)	0.018* (0.010)
Implied number of additional accidental firearm deaths	66	22	43

Figure 9: The original reported values in Table 1, which establishes a causal relationship between firearm exposure and firearm related accidental deaths.

Extension Proposals

We propose three separate extension ideas. Feedback will be very helpful in deciding the most promising direction to pursue.

Proposal 1: Study additional mass shootings

The most straight forward extension is to test whether the author’s findings hold for other national mass shootings. The given data spans 2008 - 2016 so we have ample data to test these hypotheses. Our results would discuss the generalizeability of the original paper’s findings. For example, do all mass shootings increase exposure to firearms which then leads to more accidental deaths? Looking

```
. bysort agecat: summ numdeaths pop_byage;
```

```
-> agecat = 0_14
```

Variable	Obs	Mean	Std. Dev.	Min	Max
numdeaths	12	5.5	1.930615	2	9
pop_byage	12	6.11e+07	0	6.11e+07	6.11e+07

```
-> agecat = 15p
```

Variable	Obs	Mean	Std. Dev.	Min	Max
numdeaths	12	36.41667	9.169696	25	52
pop_byage	12	2.55e+08	0	2.55e+08	2.55e+08

Figure 10: The summary statistics from Stata concur with those from the paper.

at the overall exposure trends, our initial hypothesis is that the Sandy Hook shooting produced a truly unprecedented increase in firearm purchases. If this is the case, we can dive further into characteristics that were distinct to the Sandy Hook shooting. These results could help us address the research question: which mass shootings mobilize the public to take action?

Proposal 2: Effect of gun exposure on other health outcomes

The original paper measured the effect of gun exposure on accidental deaths, which is an arguably niche outcome variable to study. Instead, as an extension, we propose studying the effect of gun exposure on more common health outcomes. One such outcome could be mental health consequences. For instance, does being exposed to more firearms affect overall mental wellbeing. The main challenge of this approach is obtaining the necessary mental health data. Second, it would be interesting to measure the effect of gun exposure on educational performance, especially as more people have advocated for giving teachers firearms to protect students. The main hurdles to this approach would be assessing the immediacy of the firearms-exposure effect (if any on education).

Nevertheless, these ideas are favorable because we already have the background checks data (firearm exposure), and we can still leverage the Sandy Hook shooting as an instrumental variable.

Proposal 3: factors effecting gun purchases

Our final proposal deviates further from the author's original objectives, but leverages their data cleaning efforts and potentially carries greater political weight. Instead of treating the number of firearm background checks as an independent variable, we are interested in using it as an outcomes variable. Presently, thanks to the original author's meticulous data cleaning, we have data on the number of background checks that occurred in each state in each month between 2008 - 2016. It would be interesting to measure the effect of various control laws on the actual number of firearms purchased. These laws could help us quantify the decrease in firearm background checks following gun control legislation. Alternatively, we could measure the potentially positive effect of media advertising (such as those from the NRA) on background checks. For this proposal, we would collect data from social media, such as Twitter, and measure the network effects driving firearm sales.

References

- [1] P. B. Levine and R. McKnight, "Firearms and accidental deaths: Evidence from the aftermath of the sandy hook school shooting," *Science*, vol. 358, no. 6368, pp. 1324–1328, 2017. [Online]. Available: <http://science.sciencemag.org/content/358/6368/1324>

Appendix

Dataverse

The original Dataverse (hosted by Harvard) contains the raw background checks, population, and trends data. It also contains the two Stata files used for data cleaning and analysis.

Project Github

We have posted all of the code and data used for replication in the private Github repository above.

Supplemental Materials

In the Google Drive folder, we have included a supplemental materials document. This file contains the Stata log-file from running the provided code as well as the R markdown document we wrote to generate the figures.