

**Fragile Families  
Gun Violence Archive  
Data on Local Deadly Gun Violence  
Restricted Use Appendage**

Year 15 Follow-Up Wave

March 2019

Funding for this Data Appendage was provided by  
The Robert Wood Johnson Foundation

***Bendheim-Thoman  
Center for Research on Child Wellbeing***  
Wallace Hall  
Princeton University  
Princeton, NJ 08544  
<http://crcw.princeton.edu>

***Columbia Population Research Center***  
1255 Amsterdam Avenue, Room 715  
Columbia University  
New York, NY 10027  
<http://cupop.columbia.edu/>

Prepared by staff in the Bendheim-Thoman Center for Research on Child Wellbeing (CRCW), Princeton University. For more information about Fragile Families, please visit our web site at <http://fragilefamilies.princeton.edu/> or email [ffdata@princeton.edu](mailto:ffdata@princeton.edu).

## DATA APPENDAGE OVERVIEW

The Gun Violence Archive Data on Local Deadly Gun Violence Restricted Use Appendage contains individual-level data on focal children's proximity to deadly gun violence incidents that occurred before the Year 15 follow-up interviews. Incident-level data on deadly gun violence incidents were collected from the Gun Violence Archive (GVA).

GVA data contain information on the precise location and date of all deadly gun violence incidents. These incident-level location data were merged to focal children's home and school addresses. The date of the deadly gun violence incidents were also linked to the caregiver's (PCG) and child's interview dates. The data appendage includes individual-level data that summarizes the number of deadly gun violence incidents that occurred within various distances (between 100 meters and 1 mile) and various time frames (between 7 days and 1 year) before the interview date. In addition, the appendage provides information on the distance (in meters) and recency (in days) of the nearest and most recent deadly gun violence incident. All information is reported in relation to the focal child's home and school location at Year 15.

## FILE LAYOUT

The file contains 4,898 observations (one per family) and is sorted by *idnum*.

## VARIABLE NAMING CONVENTION

The variable names are constructed as follows:

<u>Position</u>	<u>Character</u>	<u>Indicates</u>
1	r	Restricted data
2	g	Geographical measure
3	6	Sixth wave (Year 15)
4-7	gva_	Gun Violence Archive (GVA) data
8-12	totl_	Total GVA incidents
	recn_	Most recent GVA incident
	near_	Nearest GVA incident
13-	****	Description of time frame, proximity, and interview

For example, variable **rg6gva\_totl\_100m\_schl\_7d\_k6**: **rg6** refers to restricted data on the geographic location of the focal child at the sixth wave (Year 15); **gva\_** refers to GVA data; **totl\_100m\_schl\_** indicates this is the total number of incidents that were less than or equal to 100 meters of the child's school; and **7d\_k6** indicates that these are incidents that occurred seven days or less before the youth interview.

## MISSING VALUES

Variables containing data from the GVA have the following negative values that indicate missing data:

<u>Value</u>	<u>Label</u>	<u>Indicates</u>
-9	Not in wave	Participant did not participate in Year 15 FFCWS survey
-8	Out of range (>100km)	Any incidents in GVA data were greater than 100 kilometers from the child's home/school
-8	Out of range (>1609m)	The closest incident in GVA data was more than 1609 meters distance from the child's home/school
-8	Out of range (>365 days)	The most recent incident in GVA data was more than 365 days before the FFCWS survey date
-7	Missing lat/lon	Missing the latitude and longitude coordinates for the child's home/school

## ABOUT GUN VIOLENCE ARCHIVE DATA COLLECTION

GVA is an online archive of gun violence maintained by an independent research group and not for profit corporation ("Gun Violence Archive," 2018). GVA collects near real-time data on deadly gun violence through automated queries and manual research of internet media sources, aggregates, police blotters, and police media outlets.

GVA data is compiled at the incident-level and each incident may involve multiple shooters and multiple victims. Each incident is classified based on the location (address, latitude and longitude coordinates), date, number of people injured, number of people killed, and non-mutually exclusive categories that describe the cause and nature of the incident (e.g., suicide, murder-suicide, accidental shooting, home invasion, defensive use, domestic violence, mass shooting, officer killed subject/suspect, etc.). In the case of multiple shootings, incident-level descriptions may describe one or more of the shooters, victims, or motives related to the incident.

**For more detailed information about GVA please visit the following on-line resources:**

- GVA Home: <https://www.gunviolencearchive.org/>
- GVA General Methodology: <https://www.gunviolencearchive.org/methodology>

## DATA IN APPENDAGE

We downloaded GVA data on all gun violence incidents in the U.S. from January 1, 2014 to October 5, 2017<sup>1</sup> that resulted in one or more deaths.<sup>2</sup>

Data users should be aware that GVA does not contain data prior to January 1, 2014. We have included the variables *p6\_2014\_daysc* and *k6\_2014\_daysc* so that data users may easily determine the number of days after January 1, 2014 that the Year 15 surveys were conducted.

These data were linked to the latitude and longitude coordinates of the focal child's home and school and to the date of the youth and PCG surveys at Year 15. Home addresses were provided by the

---

<sup>1</sup> This date corresponds to the end of Year 15 FFCWS data collection.

<sup>2</sup> The data contained in this appendage were downloaded from GVA on 11/8/2017. Since GVA is continually collecting new data, it is possible that GVA may have added additional incidents following the date that FFCWS staff began preparing the data appendage.

children's PCGs at the Year 15 interview and geo-coded into geographical coordinates using Census Geocoding Services. The school attended by focal children at Year 15 was determined based on reports from children's PCGs of the name and address of the school. Geodetic distances (in meters) between adolescents' residences or schools and deadly gun violence incidents were calculated using the *geodist* command in Stata (Picard, 2010).

For this data appendage, incidents reported by GVA that involved only suicides were excluded. Incidents that involved both suicide and another form of deadly gun violence (e.g., murder-suicide, accidental shooting followed by suicide) were included.

The data appendage presents individual-level data on deadly gun violence incidents in three ways:

1. The total number of deadly gun violence incidents that occurred within various distances (between 100 meters and 1 mile) and various time frames (between 7 days and 1 year) before the interview date.
2. The distance (in meters) of the most recent deadly gun violence incident.
3. The recency (in days) of the nearest deadly gun violence incident.

Each of these is presented separately for the location of the child's home and school and for the date of the youth and PCG survey. The distance between the child's home and school is also provided (*rg6\_home\_schl\_km*). For further detail on variables included, see the Data Dictionary below.

## **COMPARING GVA DATA TO CDC DATA**

Members of our internal research team conducted an evaluation of the coverage of incidents in GVA. For this evaluation, researchers compared the number of deaths from firearms (excluding suicides) as reported by GVA to comparable statistics from the Center for Disease Control (CDC).

In total, GVA reports 12% fewer deaths between 2014 and 2016 (N=36,941) than CDC (N=42,137). For all 36 months, GVA counts are lower than CDC counts, ranging from 3% lower in June, 2014 to 18% lower in March, 2016. The difference in total counts, however, does not differ substantially across time. Compared to CDC estimates, GVA reports around 11% fewer gun deaths in 2011, 14% fewer in 2015, and 13% fewer in 2016.

GVA national counts of gun deaths closely track CDC counts over time. From 2014 to 2016, the correlation between GVA and CDC monthly counts is extremely strong ( $r=0.97$ ,  $N=36$ ). Both secular changes (annual increase in deadly gun violence) and season trends (reduction in deadly gun violence during winter months) are reflected clearly in GVA and CDC reports. For further detail on the comparison of this data, please see the Appendix below.

## DATA DICTIONARY

Variable	Description
[p k]6_2014_daysc	Number of days after January 1, 2014 (start of GVA data) the survey was conducted (PCG or youth)
rg6_home_schl_km	Distance in kilometers between home and school
rg6gva_totl_[distance]_schl_[time]_k6	# of DGV incidents within a given distance of the child's <b>school</b> for a given timeframe before the <b>youth</b> survey
rg6gva_totl_[distance]_schl_[time]_p6	# of DGV incidents within a given distance of the child's <b>school</b> for a given timeframe before the <b>PCG</b> survey
rg6gva_recn_[distance]_schl_k6	Most recent DGV incident for a given distance from the child's <b>school</b> : days before <b>youth</b> survey
rg6gva_recn_[distance]_schl_p6	Most recent DGV incident for a given distance from the child's <b>school</b> : days before <b>PCG</b> survey
rg6gva_near_[time]_k6_schl	Nearest DGV incident for a given timeframe before the <b>youth</b> survey: meters from <b>school</b>
rg6gva_near_[time]_p6_schl	Nearest DGV incident for a given timeframe before the <b>PCG</b> survey: meters from <b>school</b>
rg6gva_totl_[distance]_home_[time]_k6	# of DGV incidents within a given distance of the child's <b>home</b> for a given timeframe before the <b>youth</b> survey
rg6gva_totl_[distance]_home_[time]_p6	# of DGV incidents within a given distance of the child's <b>home</b> for a given timeframe before the <b>PCG</b> survey
rg6gva_recn_[distance]_home_k6	Most recent DGV incident for a given distance from the child's <b>home</b> : days before <b>youth</b> survey
rg6gva_recn_[distance]_home_p6	Most recent DGV incident for a given distance from the child's <b>home</b> : days before <b>PCG</b> survey
rg6gva_near_[time]_k6_home	Nearest DGV incident for a given timeframe before the <b>youth</b> survey: meters from <b>home</b>
rg6gva_near_[time]_p6_home	Nearest DGV incident for a given timeframe before the <b>PCG</b> survey: meters from <b>home</b>

### Note

Distance options include (in meters): 100, 200, 250, 300, 400, 500, 600, 700, 750, 800, 900, 1000, 1100, 1200, 1250, 1300, 1400, 1500, 1600, and 1609

Time options include (in days): 7, 14, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360, and 365

## REFERENCES

Gun Violence Archive. (2018). Retrieved September 1, 2017, from <http://www.gunviolencearchive.org>

Picard, R. (2010). GEODIST: Stata module to compute geodetic distances. Retrieved from <https://ideas.repec.org/c/boc/bocode/s457147.html>

## **Appendix: Comparing Incident-level Gun Violence Data with National Data**

Prior to releasing the Fragile Families Gun Violence Archive Data on Local Deadly Gun Violence Restricted Use Appendage, research staff in the Bendheim-Thoman Center for Research on Child Wellbeing (CRCW) at Princeton University completed a comparison of the Gun Violence Archive (GVA, <https://www.gunviolencearchive.org>) data with national data from the Centers for Disease Control (CDC).

### **CONTEXT**

No national study has combined incident-level data on violence and population-based data on individuals in the United States. This lack of research is largely due to the absence of a validated national database of violent incidents that can be geospatially linked to individual residences. Some municipalities have begun to publicly release incident-level crime data, but no harmonized national database exists to date. Other measures of neighborhood violence lack the temporal and spatial precision of incident-level crime data. For example, the Federal Bureau of Investigation's Uniform Crime Reporting data on local crime are not available for areas smaller than counties.

One promising national source of incident-level data on violence is the Gun Violence Archive (GVA), a detailed database of gun violence incidents collected from media, law enforcement, government and commercial sources. GVA has been updated in near real-time by a not-for-profit research group since 2014. Data from GVA has been used by journalists to examine patterns in gun violence across time and place. For example, The Guardian used nationwide data from 2015 to map all gun murders at the Census tract level. They found that half of all gun murders were clustered in 127 cities and that 26% of all gun murders occurred in just 1.5% of census tracts (Aufrichtig, 2017). Prior to our work, however, data from GVA has not been used in scholarly research because the accuracy and reliability of the data has not been systematically evaluated. Therefore, we tested the coverage and reliability of data from the Gun Violence Archive relative to national data on deaths by firearms available from the Centers from Disease Control (CDC) because, if validated, GVA data has the potential to be linked with population-based studies to assess individual exposure to gun violence.

### **METHODS**

#### **Incident-level data on deadly gun violence**

We obtained GVA data on all gun violence incidents in the U.S. from January 1, 2014 to October 5, 2017 that resulted in one or more deaths (N=46,701). GVA data is compiled at the incident-level and each incident may involve multiple shooters and multiple victims. Each incident is classified based on the location (address, latitude and longitude coordinates), date, number of people injured, number of people killed, and non-mutually exclusive categories that describe the cause and nature of the incident (e.g., suicide, murder-suicide, accidental shooting, home invasion, defensive use, domestic violence, mass shooting, officer killed subject/suspect, etc.). In the case of multiple shootings, incident-level descriptions may describe one or more of the shooters, victims, or motives related to the incident.

For the purposes of this comparison, incidents reported by GVA that involved only suicides were excluded from our analyses (N=2,569). Incidents that involved both suicide and another form of deadly gun violence (e.g., murder-suicide, accidental shooting followed by suicide) were included (N=2,151). Our decision to exclude suicides in this study was twofold. Conceptually, the patterns and influence of self-inflicted gun violence are likely to be different than gun violence that is directed at others. Practically, underreporting of suicides in official statistics is substantial (Tøllefsen, Hem, & Ekeberg, 2012). We anticipated such underreporting to be especially great in non-official data sources like GVA that predominantly rely on news reports.

## **Aggregate-level data on deadly gun violence**

To evaluate the coverage of incidents in the GVA, we compared the number of deaths from firearms (excluding suicides) as reported by GVA to comparable statistics from the CDC. We aggregated the total number of non-suicidal deaths per incident in GVA to national and state counts by month and year from January 2014 to December 2016. We assumed zero non-suicidal deaths when GVA reported one death and classified the incident as suicide. We assumed one non-suicidal death when GVA reported two deaths and classified the incident as murder-suicide. For multiple death incidents classified as suicide (two or more deaths) or murder-suicide (three or more deaths), we determined the number of non-suicidal deaths by manually reviewing the original source cited by GVA (e.g., local newspaper or police blotter).

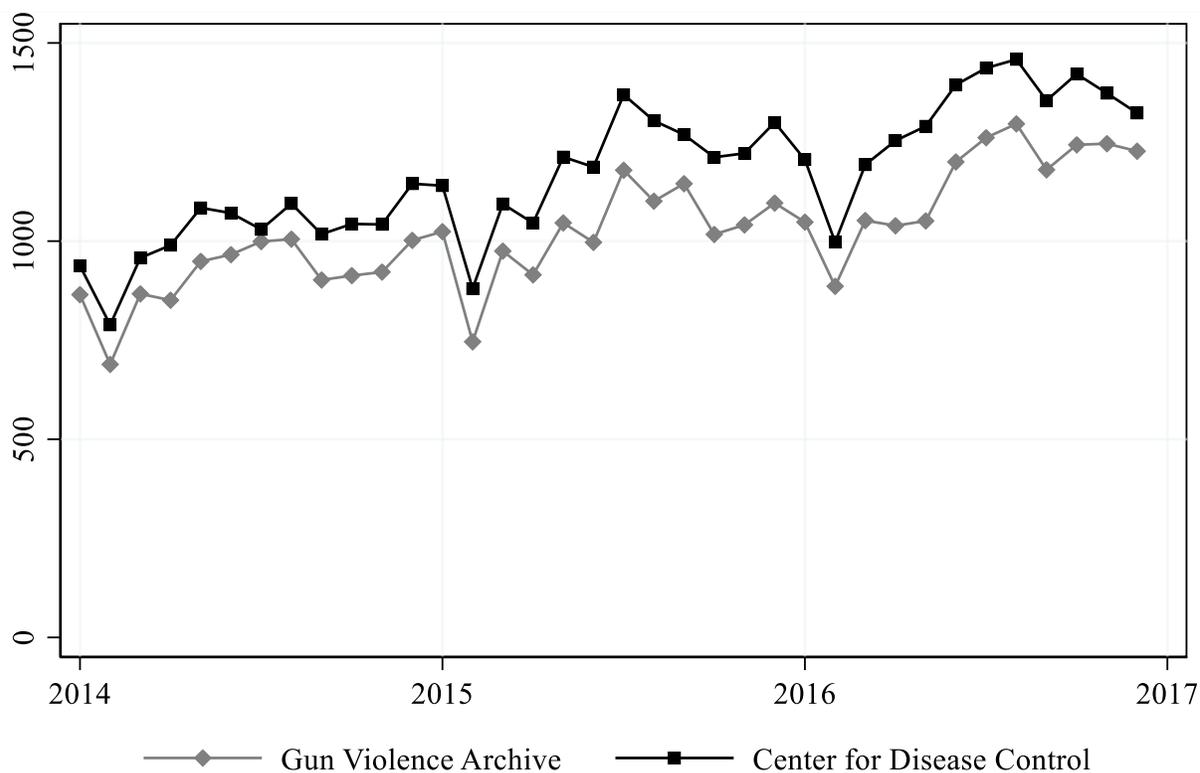
Aggregate-level data on deadly gun violence was downloaded from CDC's Underlying Cause of Death data by state and month. CDC mortality data are based on all death certificates filed by all U.S. states and the District of Columbia. Mortality data compiled by states are then provided to the National Center for Health Statistics for aggregation and reporting through the Vital Statistics Cooperative Program. We used CDC's Wide-ranging OnLine Data for Epidemiologic Research (WONDER) to filter mortality data by injury intent and mechanism, including all deaths caused by firearms for which the injury intent was not classified as suicide (i.e., unintentional, homicide, undetermined, legal intervention, and no intent classified).

## **RESULTS**

### **Comparison of GVA and CDC reports of gun deaths**

We evaluate the coverage and reliability of GVA data by comparing GVA reports of non-suicidal gun deaths to corresponding statistics from the CDC. We focus on differences in and correlations between monthly counts over time. Figure 1 shows the number of gun deaths per month in the U.S. between January, 2014 and December, 2016 as reported by the two data sources. In total, GVA reports 12% fewer deaths between 2014 and 2016 (N=36,941) than CDC (N=42,137). For all 36 months, GVA counts are lower than CDC counts, ranging from 3% lower in June, 2014 to 18% lower in March, 2016. The difference in total counts, however, does not differ substantially across time. Compared to CDC estimates, GVA reports around 11% fewer gun deaths in 2011, 14% fewer in 2015, and 13% fewer in 2016.

### U.S. Gun Deaths per Month (Excluding Suicides)



**Figure 1.** U.S. Gun Deaths per Month (Excluding Suicides) as Reported by Gun Violence Archive and Center for Disease Control.

From 2014 to 2016, the correlation between GVA and CDC monthly counts is extremely strong ( $r=0.97$ ,  $N=36$ ). Both secular changes (annual increase in deadly gun violence) and season trends (reduction in deadly gun violence during winter months) are reflected clearly in both GVA and CDC data.

Table 1 shows the number of gun deaths (excluding suicides) by state for the 25 states with the most CDC reported gun deaths – around 88% of all U.S. gun deaths during this period. Because CDC suppresses counts below 10, monthly counts are not consistently available for the majority of the remaining states. For states shown in Table 1, CDC reports more gun deaths than GVA (12% on average, like for the U.S. as a whole). There is noteworthy state-level variation in the difference between counts. In Michigan, for example, GVA reports 32% fewer gun deaths than CDC. In Wisconsin, South Carolina, and Kentucky, however, the difference is only 2%.

**Table 1. Gun Deaths (Excluding Suicides) between 2014 and 2016 by State**

	Gun Violence Archive	Center for Disease Control	% Difference
Alabama	1104	1262	-13%
Arizona	634	846	-25%
Arkansas	470	561	-16%
California	3526	4485	-21%
Florida	2344	2833	-17%
Georgia	1501	1875	-20%
Illinois	2157	2355	-8%
Indiana	973	1025	-5%
Kentucky	624	636	-2%
Louisiana	1377	1483	-7%
Maryland	1050	1189	-12%
Michigan	959	1409	-32%
Mississippi	682	912	-25%
Missouri	1279	1338	-4%
New Jersey	769	869	-12%
New York	1152	1239	-7%
North Carolina	1430	1569	-9%
Ohio	1490	1601	-7%
Oklahoma	586	663	-12%
Pennsylvania	1437	1625	-12%
South Carolina	1012	1035	-2%
Tennessee	1114	1282	-13%
Texas	3127	3605	-13%
Virginia	887	957	-7%
Wisconsin	524	533	-2%

Note: Table shows the 25 U.S. states with the largest number of gun deaths as reported by CDC between 2014 and 2016.

Finally, Figure 2 compares monthly counts by state over time. In general, counts within each state are strongly correlated: the average within-state correlation is  $r=.86$ . California is the only state where GVA and CDC counts are only modestly correlated ( $r=.46$ ). Much of the divergence between GVA and CDC counts within California appear to have started in 2015 when CDC counts increased substantially, but GVA counts did not. We highlight this exception given the state population and relative contribution to total gun violence in the U.S.: California alone accounts for around 12% of the U.S. population and 10.6% of all gun deaths reported by CDC over this period (9.5% reported by GVA).

## U.S. Gun Deaths per Month (Excluding Suicides) by State



**Figure 2.** Gun Deaths per Month (Excluding Suicides) by State as Reported by Gun Violence Archive and Center for Disease Control. The figure shows the 25 U.S. states with the largest number of gun deaths as reported by CDC between 2014 and 2016.

## DISCUSSION

Overall, these findings point to adequate GVA coverage and reliability of deadly gun violence in the U.S. Aggregate GVA data on deaths from gun violence (excluding suicides) very closely track CDC data over time. Assuming perfect reporting by CDC, GVA undercounts around 12% of non-suicidal gun deaths. This modest undercount does not vary substantially across time (between 2014 and 2016), but does vary modestly across states.

## REFERENCES

- Aufrichtig, A. (2017). Mapping US gun murders at a micro level: new data zooms in on violence. Retrieved September 1, 2017, from <https://www.theguardian.com/world/2017/mar/20/mapping-gun-murders-micro-level-new-data-2015>
- Gun Violence Archive. (2018). Retrieved September 1, 2017, from <http://www.gunviolencearchive.org>
- Tøllefsen, I. M., Hem, E., & Ekeberg, Ø. (2012). The reliability of suicide statistics: a systematic review. *BMC Psychiatry, 12*(1), 9.