Firearm Acquisition-Related Online Search and Gun Violence across the COVID Purchasing Cycle, 2015-2025

Abstract

Objectives. Firearm purchasing in the United States surged in 2020 and declined slowly. This study examines whether fatal gun violence rose after Google searches about acquiring guns increased, across 210 US media markets from 2015 to 2025.

Methods. Because no national record of gun purchases exists, and because earlier studies found that gun-related Google searches track firearm background checks, we built a monthly index of searches about buying, storing, and maintaining guns (acquisition-related search) for every media market. Within states, our Google Search index also tracks background checks. In population-weighted fixed-effects regressions, we related the index to fatal shootings recorded by the Gun Violence Archive within the following year, comparing each market with itself over time. We repeated the analysis with the FBI's Supplementary Homicide Reports and CDC death certificates.

Results. Increases in acquisition-related search were followed within a year by increases in fatal gun violence. At a six-month lag, the association appeared in all three data sources, including state-level death certificates, and the twelve-month association held after the seasonal pattern of search was removed. As expected, the association appeared only for firearm homicide, not for non-firearm homicide. However, the association was strongest during the 2020-2021 surge and came mostly from roughly the fifty largest markets, home to two-thirds of Americans.

Conclusions. From 2020 onward, acquisition-related online search was associated with fatal gun violence within the following year, most strongly during the 2020-2021 surge. Limitations include the use of search as a proxy for purchases and comparisons that rest on regional differences in timing.

Keywords: Google Trends; firearm acquisition; gun violence; measurement validation; panel data

1. Introduction

In 2020, the United States saw a sharp increase in firearm homicide. Kegler et al. (2022) reported that the age-adjusted rate grew by nearly 35 percent within one year, to its highest level since 1994. Around the same time, firearm purchasing went through a complete cycle, with a sharp surge that started in March 2020 and a slow decline afterward (Fig. 1). Several studies have asked whether the two rises are connected. According to Degli Esposti et al. (2025), background checks had begun to rise by mid-2019 and firearm homicide by October 2019, and the authors listed the surge in purchases as one possible mediator of the firearm-specific rise. With a single national series for each, however, the timing of one rise cannot separate new guns from other changes in the same months. Tests with state and city data found little link between gun purchasing or availability and violence (Schleimer et al. 2021; Fazel-Zarandi and Barnett 2024).

One reason the earlier tests found little may be that the association appears only with a delay. Cook (2018) studied gun markets through the transactions that arm offenders, because offenders usually acquire a gun within weeks or months of using it, and he described the guns already in private hands as the reservoir from which most of these transactions draw. If some guns bought during a purchasing surge are misused by their buyers or pass to offenders, gun violence should rise in the months after the surge, and it should rise first where the surge came earlier. Comparing places where interest in buying guns rose at different times can test this prediction, because changes that hit the whole country at once drop out of the comparison.

Testing the prediction above requires firearm purchasing data for places smaller than the nation, and such data are scarce. There is no national firearm registry, and the basic information about gun acquisition was lacking (National Research Council and Institute of Medicine 2013). Federal law bars the Attorney General from requiring, by rule or regulation, any system of registration of firearms, firearms owners, or firearms transactions (Firearms Owners' Protection Act, 18 U.S.C. § 926(a)). A permanent appropriations rider, known as the Tiahrt Amendment, also bars the disclosure of record-level trace data outside law enforcement, although aggregate statistics may be published (Consolidated Appropriations Act, 2010). State point-of-sale records, such as the California Dealer's Record of Sale used by Wright et al. (2010), cover only their own state. The closest national record is the count of background checks run through the National Instant Criminal Background Check System (NICS). As Brownstein et al. (2020) summarized, these checks are not purchase counts. The checks are also published only as monthly totals for each

state. As a result, no national measure of purchase flows is available in the United States, let alone subnational estimates. Search data have been related to firearm outcomes at the national or state level. Brownstein et al. (2020) found that states with more searches for some firearm terms had higher firearm death rates, and Johnson et al. (2026) found that state gun homicides did little to move gun-related search before 2020. Most earlier studies relied on national trends or state data, whereas we compare media markets, below the state level.

We ask whether fatal gun violence rose after online searches about acquiring guns increased. In this paper, we work with an index of acquisition-related search, a fixed basket of seven Google search phrases about buying, storing, and maintaining guns. We built the index for all 210 media markets of the United States, month by month from 2015 to 2025. The platform returns a slightly different value each time the same data are requested. For this reason, we retrieved every market-month thirty times, and we used five anchor markets to put all months on one scale. Within states, the index tracks background checks. We find that increases in acquisition-related search were followed within a year by increases in fatal gun violence, and the response held after the seasonal pattern of search was removed. At a six-month lag, the association appears in two other sources of homicide data, and it appears in firearm homicide but not in non-firearm homicide. It was strongest during the 2020-2021 surge, was not detected before 2020, and comes mostly from roughly the fifty largest markets, home to about two-thirds of the US population. It is regionally identified, which means that the estimate rests on differences in wave timing across regions and not on variation within states.

To our knowledge, no earlier study has related a monthly, search-based measure of gun acquisition below the state level to fatal gun violence. Section 2 reviews research on guns, purchasing, and violence and on the use of search data in social science and in research on crime and guns, and it sets out the gaps that this study addresses. Sections 3 and 4 describe the data and the research design, Section 5 reports the results, and Section 6 discusses our findings and their limits.

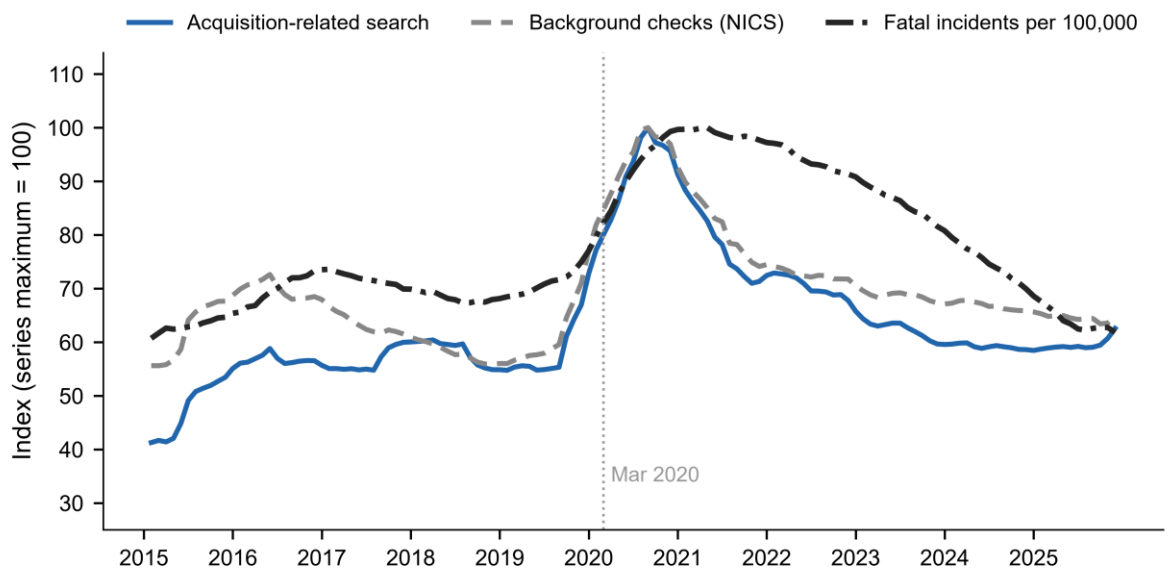


Fig. 1 National trends, 2015-2025, in acquisition-related search (calibrated and population-weighted), NICS background checks (handgun, long-gun, and multiple-check categories), and fatal interpersonal incidents per 100,000 residents (GVA). The vertical dotted line marks March 2020. Each series was smoothed with a centered 12-month mean (with a shorter window at the start and end of the period) and indexed to its own maximum. Because the mean is centered, each smoothed line starts to move about five months before a change in the monthly data.

2. Background

2.1 Guns, purchasing, and violence

Research on guns and violence finds that more firearm homicide is usually associated with more gun ownership. Duggan (2001) used sales of the magazine *Guns & Ammo* as a proxy for ownership, and Cook and Ludwig (2006) used the share of suicides committed with a firearm. In both studies, the area-level association was concentrated in gun homicide. Across states from 1981 to 2010, each percentage point of household gun ownership went with a 0.9 percent higher firearm homicide rate (Siegel et al. 2013). At the household level, Kellermann et al. (1993) found in a case-control study that a gun in the home was associated with a higher risk of homicide in the home. In a California cohort, adults who lived with handgun owners died by homicide at more than twice the rate of those who did not (Studdert et al. 2022).

However, it is still disputed whether the stock of guns causes violence. Kovandzic et al. (2013) showed that the stock proxies barely track direct measures of ownership over time within areas. In a county cross-section from 1990, they estimated a negative effect of gun prevalence on homicide, which they attributed mainly to the deterrent effect of guns held by noncriminals. Kleck (2015) concluded that none of the studies meeting all three of his methodological criteria supported the hypothesis that higher gun ownership raises crime rates, and he argued that lagging a predictor evades the problem of causal order instead of solving it. Hayo et al. (2019) argued that the county results of Cook and Ludwig (2006) were spurious, because the outcome and the explanatory variables were divided by the same population, and the association disappeared once population was entered as a control. Cook and Ludwig (2019) replied that their county estimates did weaken under this specification but that their state estimates did not. Morral et al. (2024) used a cross-lagged design and found that differences in household ownership were followed by differences in firearm suicide, but not significantly by differences in firearm homicide. The stock also moves slowly. Kleck (2015) noted that the share of households reporting a gun rarely fell outside the range of 44 to 52 percent from 1959 to 1993, which may have left longitudinal studies little change to estimate from.

Purchases measure something different from the stock. Cook and Ludwig (2006) pointed out that a purchase by a household that already owns a gun does not change prevalence. Azrael et al. (2017) estimated that about seventy million firearms changed hands in 2011-2015, while the share of adults who own guns changed little. The guns used in violence usually reach offenders shortly before they are used. In a survey of Chicago inmates, Cook et al. (2019) found that the median time from acquiring the gun to the crime was two months. Most of these guns come from the existing private stock (Cook 2018), but some new guns move from retail sale to crime quickly. Wright et al. (2010) followed handguns sold in California in 1996 and found that 0.4 percent were recovered by police from someone other than the buyer within three years. Braga et al. (2012) reported that new handguns are overrepresented among recovered crime guns, which points to diversion close to retail. According to the Department of Justice, nearly one in four traced crime guns in ATF data for 2017-2021 had been sold at retail less than a year before its recovery, although fewer than half had been sold within three years (US Department of Justice 2023). Of the 604,968 traced firearms that police recovered within a year of retail sale in 2015-2023, 83.3 percent were recovered in the state where they had been sold (Weigend Vargas and

Goldstick 2026). In Chicago, Cook et al. (2007) described the underground market as thin, and they linked the shortage of willing sellers in part to low household gun ownership. A rise in legal ownership could therefore make sellers easier to find. Overall, some new guns reach crime within months, and the trace data place most of them in the state of sale.

Studies of purchase rules and of short purchase spikes give mixed results for violence. Luca et al. (2017) estimated that waiting periods, which delay a purchase by a few days, reduce gun homicides by roughly 17 percent. Edwards et al. (2018), on the other hand, found no significant change in homicide from purchase delays. Koenig and Schindler (2023) studied the six-month demand shock of 2012-2013 and found that states with purchase delays had smaller increases in handgun sales and 2 percent lower homicide rates. Webster et al. (2014) found that Missouri's 2007 repeal of its handgun purchaser licensing law was associated with an increase in firearm homicide but not in non-firearm homicide. Short purchase spikes have been followed by more firearm harm, mostly accidents and injuries. Levine and McKnight (2017) estimated that about three million additional guns were sold after the Sandy Hook shooting. Accidental firearm deaths rose over the following five months, but homicide did not respond. In California cities, Laqueur et al. (2019) found that larger handgun spikes after the 2012 election and the Sandy Hook shooting were followed by more firearm injuries over the next year, although they found no significant association for firearm deaths or for shorter windows. Studdert et al. (2017) found that the handgun spikes in California after the Sandy Hook and San Bernardino shootings were short-lived and made up less than 10 percent of annual acquisitions.

The surge of 2020-2021 was larger than these spikes, and purchasing stayed high for almost two years (Fig. 1). Miller et al. (2022) estimated that 7.5 million adults became new gun owners between January 2019 and April 2021. Between 2018 and 2022, firearm homicide rose by 54 percent while non-firearm homicide remained stable (Degli Esposti et al. 2025), and crime guns in California were recovered sooner after their sale, as the share of handguns recovered in violent crimes within a year of their last recorded sale rose from 7 percent in 2010 to 33 percent in 2021 (Laqueur et al. 2023). However, the direct tests of the surge found little association between purchasing and violence. Schleimer et al. (2021), who used state data through July 2020, found no association between excess purchasing and non-domestic firearm violence after adjusting for pandemic conditions, and Fazel-Zarandi and Barnett (2024) found that, across 50 large cities,

homicide growth in 2020 was not concentrated in the cities with the greatest availability of new and older guns.

Purchases were not the only change in these years that could have raised firearm violence. As Degli Esposti et al. (2025) noted, accounts of the rise have often pointed to the pandemic, with the disruptions and stressors it brought, and to the unrest that followed the police killing of George Floyd in May 2020. In national weekly data, the same authors found that 28 percent of the eventual increase in firearm homicide had occurred by the time COVID-19 was declared a national emergency, and 65 percent by the time of George Floyd's death. The authors concluded that these events may have steepened a rise that had already begun, and they noted that the rise also coincided with the campaign for the 2020 national election. Because they may also have reached places at different times, a comparison across places has to account for them, and Section 4 describes how we do so.

The direct tests leave the timing of the surge below the state level unexamined. Purchases are recorded nationally only as state-month background checks, which, as Schleimer et al. (2021) noted, are not available below the state level. The existing tests therefore use states, the cities of one state, short windows after single events, or a single cross-section of cities. Schleimer et al. (2021) called for research on the longer-term risks of the surge, and Sharkey et al. (2024), who found more shootings in rural counties in the first two weeks of deer season, called for studies of sudden changes in firearm availability in densely populated places.

2.2 Google Search data in social science

Researchers first used search data as a timely record of outcomes that official statistics report with a delay. For example, Ginsberg et al. (2009) estimated weekly influenza activity in each region of the United States from search queries, Askitas and Zimmermann (2009) related keyword searches to unemployment in monthly German data, and Choi and Varian (2012) used search to forecast economic indicators such as automobile sales and unemployment claims. Social scientists later used search data to measure attention to political issues (Mellon 2014) and attitudes that people hide from surveys (Stephens-Davidowitz 2014).

Search data add the most where official data are missing. Goel et al. (2010) found that, for tracking influenza, search improved only modestly on a simple autoregressive model, and they concluded that search queries provide a useful guide in the absence of other data sources or

where small gains in accuracy matter. Google Flu Trends shows what can go wrong when search is used in place of an official series that already exists. It was built to predict the influenza reports of the Centers for Disease Control and Prevention (CDC), and Lazer et al. (2014) found two problems that contributed to its errors. The search terms of its first version had been chosen because they fit past flu data, and the model assumed that the link between search volume and flu stayed fixed while Google kept changing its search service. We share their view that big data should supplement traditional measurement and should not substitute for it. They also argued that measuring flu at very local levels, which the CDC cannot practically do on a wide scale, would be worth more than improving national estimates. A reassessment of Google Flu Trends, however, found flaws at the national, regional, and city scales, and in most periods the errors in epidemic intensity were largest in New York City (Olson et al. 2013). A measure below the state level therefore needs validation of its own.

Work on the methods of search data shows where such measures go wrong and how the errors can be reduced. Eichenauer et al. (2022) showed that, because the platform samples searches, identical requests return different values, and that this noise is largest for less popular terms, smaller regions, and shorter time units. From one unplanned experiment, Franzén (2023) concluded that scientists should refrain from using Google Trends, while Raubenheimer (2024) showed that the variation is controlled when several samples are obtained and aggregated. The recommended remedy is therefore to average repeated samples (Hölzl et al. 2025). However, averaging does not remove the errors that Google's own changes in data collection introduce. Another limitation is that Google Trends updates can change what the same query returns over time (Liu 2024).

Mellon (2014) set out a general method for validating such measures. The search terms should look like the concept they are meant to measure (face validity), the searches should reflect that concept and not other meanings of the same words (content validity), and the index should move with an existing measure of the same concept once seasonal patterns and trends are taken out (criterion validity). Few applications meet even basic checks. In the review by Hölzl et al. (2025), about a third of the 360 studies ran any validity check, and seven percent drew more than one sample. The authors also noted that the methods proposed for putting regions and months on one scale had not been thoroughly examined, and that Google Trends itself offers no full panel output. Even so, they saw in Google Trends the potential to complement surveys and

administrative records, provided that researchers address these problems, for example by validating their keywords and averaging several samples. In criminology, Johnson et al. (2026) recently used Google searches and background checks to study gun demand in the states, and they checked their results with placebo searches and with permit data from Tennessee. We report the corresponding checks for the present index in Section 3.1 and in Supplements A.7 and A.8.

2.3 Google Search data in research on crime and guns

In research on crime, search data have mainly been used to measure what official records miss. During the COVID-19 lockdown in Greater London, Anderberg et al. (2022) observed that their search-based index of domestic violence rose by 40 percent at its peak, seven to eight times the increase in police-recorded crimes. Berniell and Facchini (2021) reported a 30 percent rise in searches for topics related to domestic violence after lockdowns in eleven countries. Liu et al. (2023) found that search-based estimates of motor vehicle theft, larceny, and rape were reliable across 107 media markets, and they showed how such estimates can identify places where many offenses go unreported.

Most of these studies use search as the outcome or as the measure being tested. Bestenbostel and Peralta (2026), for example, reported that racist internet searches and hate crimes rose after notices of mass layoffs, using a county-month panel with searches measured by media market. In research on crime, the closest earlier use of search as an explanatory variable in a monthly panel of media markets is Lu and Sheng (2022). As a check on their measure, they related a monthly search index for 60 media markets to monthly anti-Asian hate crimes. Their index is scaled to each market's own peak, which, in their words, captures "the timing but not the level" of a change.

Firearm research has used search mainly to show that interest in guns rose after the Sandy Hook shooting and at the start of the pandemic, and that it moves with background checks. Levine and McKnight (2017) found that weekly national searches for "buy gun" and "clean gun" spiked immediately after the Sandy Hook shooting. In the first weeks of the COVID-19 pandemic, searches related to buying and cleaning guns were 158 percent above the level expected from earlier trends (Caputi et al. 2020). Brownstein et al. (2020) correlated single search terms with background checks, firearm deaths, and state firearm policies, using national monthly data for 2019 and cross-sections of the 50 states.

Gun-related search has also stood in for gun ownership and served as an outcome. In Stephens-Davidowitz (2014), the search rate for "gun" explained 62 percent of the variation in state gun ownership rates. Johnson et al. (2026) asked whether gun demand in each state follows local conditions or national ones. With monthly searches containing the word "gun" and background checks for the 50 states, mainly for 2004-2019, they found that the states moved closely together in both series and that state gun homicides had little effect on search on average before 2020, and only a moderately different effect in data centered on 2020. State conditions mattered mainly during periods of extreme national stress. They drew their search data from Google's Health Trends interface, which returns search shares without the usual rescaling, and they used the broad term "gun" partly because narrower phrases are much noisier in smaller areas. In their study, homicide enters as one possible local driver of search. We ask the opposite question, whether fatal violence followed the demand, and we measure search for each media market instead of each state. Their main analysis covers 2004-2019 and extends to the years around 2020, while our main estimates run through 2025 and include the 2020-2021 surge.

2.4 The current study

The literature reviewed above leaves two gaps that this study addresses. The first gap concerns measurement. Gun purchasing is not recorded below the state level in any national source, and public search data are difficult to use at that level, as each request is scaled separately and narrow search phrases are noisy in small areas (Eichenauer et al. 2022; Johnson et al. 2026). To fill this gap, we constructed a monthly index of acquisition-related search for all 210 media markets from repeated retrievals and placed all months on one scale with anchor markets. We measured its reliability across the retrievals and checked its validity against state background checks and an independent retrieval collected four years earlier (Section 3.1).

The second gap concerns testing. The earlier tests reviewed in Section 2.1 did not follow a purchase wave month by month below the state level. We follow the 210 media markets from 2015 to 2025, through the rise and the retreat of the surge, and ask whether fatal gun violence followed acquisition-related search within the same market. We use differences across markets in when the wave arrived, compare three sources of data on fatal violence, and test whether the association is specific to firearms.

3. Data and Measurement

3.1 The GT search panel

The query set is a basket of seven acquisition-related phrases: "how buy gun", "where buy gun", "gun sales", "gun safe", "gun maintenance", "gun shop near me", and "gun store near me". Together, they span buying, storing, and maintaining guns. Searches about gun maintenance have been tracked before, by Levine and McKnight (2017), who followed "clean gun" together with "buy gun". We submit all seven phrases as a combined query, so the platform returns one index for the whole basket. We avoid single terms because they carry too little volume to return meaningful data. On the other hand, the platform also provides a "topic" entity, which collects search terms on the same concept across languages. Despite returning a high search volume, it aggregates all gun-related searches regardless of user intent. Consequently, users searching for gun issues or gun violence discussions are mixed with those searching for how to buy, own, or store a firearm. Our approach of combining the chosen terms keeps the acquisition intent while pooling enough volume for smaller markets to return values. A content-validity check is in Supplement A.8. It uses the platform's related queries to check that these phrases capture the acquisition intent, following Mellon (2014).

For the seven acquisition-related phrases combined, Google Trends (GT) reports an index of search activity across US designated market areas (DMAs, hereafter media markets or markets). These markets are the finest geographic unit in the Google Trends platform that covers the search data from the whole US. We use the month as the time unit because over shorter windows, such as days, search volume in most markets falls below the platform's reporting threshold and returns zero. The unit of analysis of this study is therefore the market-month. We choose our data range from January 2015 to December 2025, roughly five years on either side of the March 2020 COVID-19 lockdowns, so that the purchasing surge, its retreat, and the full seasonal cycle all fall inside the window.

If we pull the data by geographic area in a single month, the platform scales every market from zero to one hundred against that month's highest-share market, so we can compare search shares across markets in that same month. However, data pulled this way are not comparable across months. A market at sixty in January and sixty in February has not necessarily held the same search share, because each sixty is measured against a different month's top market. To transform

these cross-sectional monthly data into a comparable panel, we need a scale that holds across months.

The monthly data can be joined with the help of a second way of pulling the data. Splicing fragmented Trends downloads is an established problem with published solutions (Eichenauer et al. 2022; Bleher and Dimpfl 2022), and ours follows the anchor-bank logic of West (2020). Querying one single geographic area over the full time range returns that area's zero-to-one-hundred scale across the whole range, consistent over time within that area. A series pulled this way can anchor the monthly cross-sections. We retrieved full-period series of this kind for five markets and five states. Anchors were chosen after the first retrieval round on the condition that search volume is high enough to clear the reporting threshold in every one of the 132 months, so each anchor's ratio series has no gaps. The five markets span the size range from New York down to a mid-sized market, and Supplement A.2 details the selection. The monthly scale factor is computed from the anchors as

$$scale(t) = \text{median}_b\left\{\frac{T_b(t)/I_b(t)}{\text{mean}_t\,[\,T_b(t)/I_b(t)\,]}\right\} \qquad (1)$$

where $T_b(t)$ is anchor b's directly retrieved full-period series, $I_b(t)$ is its value in month t's cross-market table, $\text{mean}_t[\cdot]$ averages the ratio over the 132 months, and median_b takes the median over the five anchors. Both $T_b(t)$ and $I_b(t)$ carry anchor b's true search share, so the share drops out of the ratio $T_b(t)/I_b(t)$. What is left is the top market's share ($s_{\text{top}}(t)$) that month over the anchor's own peak-month share ($s_b(t^*)$). Dividing the ratio series by its own average over the months, the denominator in Eq. (1), removes that peak-month share ($s_b(t^*)$) as well, so what survives depends only on the month ($s_{\text{top}}(t)$).

Eq. (1) gives the top market's share in each month as a ratio to the average month, and $scale(t)$ carries the month-to-month movement of that share. The monthly cross-market table lacks exactly this piece. The table reports each market's share divided by that month's top-market share, and this divisor changes every month. We multiply the table by $scale(t)$ to replace the changing divisor. The calibrated value is therefore each market's true share times one common constant. Supplement A.2 walks through the arithmetic with a worked example.

The robustness test leaves out any one of the five anchors, calibrates with the other four, and predicts the left-out anchor's own directly retrieved series; it gives r = .979 to .9996 (Supplement

A.3). Also, our main conclusion from the fixed-effects models does not depend on the scaling, because the search measure enters in logs, the scale factor becomes a month fixed effect, and the estimate is the same with or without calibration (Supplement A.4). We use the calibration for the panel itself, to make the data comparable across months and regions, and for the national comparisons.

Another issue in the Google Trends data is that the data itself shifts slightly over time, because the data the platform returns come from a sample of searches. Retrieving the same market-month again after a few hours returns different values. We therefore sampled every market-month thirty times with a Python scraper driving a full Chrome browser (Selenium; Supplement A.1 details the pipeline). Passes were spaced at least four hours apart to make each scrape likely to draw a new sample from the platform.

All main-panel retrievals were collected in July 2026. Averaging the data suppresses sampling noise, and Supplement A (Table A1) reports the reliability of the averaged panel, a Cronbach alpha of .97 for markets and .997 for states (Cronbach 1951). Averaging across passes also recovers small markets, which fall below the reporting threshold in some passes but not in others. The averaged panel of thirty samples holds more market-months than any single retrieval does. The cells that stay empty in every pass are handled as described in Section 3.4.

We independently checked the data's validity with administrative data and with a previous scrape of GT data from the same phrases collected in 2022 (roughly 200 samples). We first aggregated the market panel to states and compared it with the state panel retrieved directly from the platform. The two series match at a median within-state correlation of .98 across 48 states. We then compared GT search data with NICS background checks (cleaned to handgun, long-gun, and multiple-check categories). The state panel tracks the NICS checks at a median within-state correlation of .78 across 50 states. The market panel aggregated to states tracks them at .77 across 47 states. By period, across the 49 states other than Hawaii and Kentucky, the median within-state correlation is .71 in 2015-2019, .82 in 2020-2021, and .67 in 2022-2025. Part of this agreement is seasonal. With calendar-month means removed from both series, the correlation stays at .83 in 2020-2021 but falls to .48 in 2015-2019 and .55 in 2022-2025. The slope of the link also varied, as the state elasticity of background checks with respect to search was .19 in 2015-2019, against .43 in 2020-2021 and .37 in 2022-2025 (Supplement A.7). The 2022

replication scrape correlates with the present GT panel at a median monthly cross-sectional correlation of .887 across states and .639 across markets (Supplement A.7).

3.2 The gun violence data

The main firearm violence data came from the Gun Violence Archive (GVA), which compiles incidents from media reports and from local and state police sources (Gun Violence Archive 2026b). GVA provided the incident-level file upon inquiry. Each record carries the date, the location, the numbers killed and injured, and the characteristics of the incident, such as accidental, officer-involved, and child-involved shootings (Gun Violence Archive 2026a). From 2015 to 2025, after dropping cases without coordinates, accidental shootings (including those in which a child shot themselves or another person), and suicides or suicide attempts with no other victim (murder-suicides are kept), the number of incidents has been cut from 433,988 to 407,323. There are 157,793 fatal incidents, and the other 249,530 are nonfatal. Supplement C lists each tag rule and the count remaining after each step. These are then merged into counties by their coordinates and into media markets through the county-to-market crosswalk (Sood 2016). Our main analysis uses fatal incidents from GVA, in order to align with the other data sources introduced below, which provide fatal shooting data only.

Two other sources serve as robustness checks for the fatal shooting data. The National Vital Statistics System (NVSS) is available at the state-month level, and it codes the cause of death from death certificates. We count deaths whose underlying cause is assault by firearm discharge, ICD-10 codes X93 to X95. The 2015-2025 NVSS data come from three CDC WONDER databases spliced by year, final data for 2015-2020 and for 2021-2024 and provisional data for 2025 (Centers for Disease Control and Prevention 2021, 2026a, 2026b); the databases match on their overlapping years ($r = 1.00$). The Supplementary Homicide Reports (SHR) is from the FBI's Uniform Crime Reporting Program. It records the weapon used in each homicide reported to the FBI. SHR data are available at the county level and are aggregated to markets through the same county-to-market crosswalk as GVA. We obtained the SHR file from Kaplan (2025). The homicide reports are filed voluntarily. In 2021, after the FBI retired its older summary reporting system (Federal Bureau of Investigation 2021), the reports for several of the largest markets fell far below their 2020 and 2022 levels, so we leave that year out of the homicide-report estimates (Supplement B.4). Florida is also left out, because the program covers it inadequately. In 17.6 percent of the remaining market-months the reports record no homicide, which cannot be told

apart from a month in which no agency filed, so their coverage is not random. We validate GVA against both NVSS and SHR. Across the 3,904 state-months of the 32 states with at least 60 unsuppressed months, GVA death counts and NVSS firearm-homicide counts correlate at r = .98 (Supplement A, Table A1). At the market-month level, the GVA fatal-incident rate and the SHR firearm-homicide rate correlate at r = .72, and the median correlation within markets is .75 (Supplement A, Table A1).

3.3 Control and other auxiliary data

Population data are from Census Bureau county estimates, aggregated to markets and interpolated to months from the yearly estimates (US Census Bureau 2020, 2026). The population data for the final five months of 2025 are extrapolated, because no estimate for them had been released when the data were collected. The Oxford COVID-19 Government Response Tracker stringency index (Hale et al. 2021) measures how strict each state's government restrictions were during the pandemic. Google county mobility data (Google LLC 2022), based on anonymized phone location records, give the percent change in visits to workplaces, homes, and retail places compared with the weeks before the pandemic. The Crowd Counting Consortium provides protest event counts for 2017 to 2024 (Chenoweth et al. 2025a, 2025b). State firearm-law stocks come from the RAND State Firearm Law Database, compiled through September 2025 (RAND Corporation 2026). Four other series of data were collected for our robustness and discussion sections: county unemployment (US Bureau of Labor Statistics 2026), county temperature and precipitation (Vose et al. 2014), officer-involved shootings from GVA's own incident tags (Gun Violence Archive 2026a), and state-month arrest counts for weapons, drug, and all offenses (Federal Bureau of Investigation 2026).

3.4 Missing data

The search panel does not cover every market-month. Google reports a value for a market only when its search volume in that month is above the platform's threshold, so a missing value means that searches were too low to be reported, not that the retrieval failed. A market-month is missing only if all thirty retrievals returned nothing, which is the case for 2,146 of the 27,720 market-months (7.7 percent), mostly in small markets and in the early years (Supplement A.6; Fig. A1). Weighted by population, the missing share is 0.9 percent.

We estimate the models on the observed market-months of all 210 markets. A market-month goes missing when the market is small or its search that month was low. The market fixed effects absorb stable differences between markets, and the population weights give small markets little weight. If missingness depended on violence, an indicator for whether the previous month was observed would predict the outcome. In Supplement D.1 it does not, and adding it leaves the search coefficient unchanged. The estimates therefore describe the market-months in which search was high enough to be reported.

We do not impute the missing values. Interpolation, median imputation, and multiple imputation all assume that a missing month resembles the observed months of its market, but a month is missing precisely when its search volume is lower than in the observed months. Imputation of this kind therefore fills the gaps with values that are too high. Restricting the sample to markets with more complete search data changes the estimate little or raises it, while every method of filling the missing cells lowers it (Supplement D.1).

4. Research Design

In order to describe the timing of the association, we start with one lag at a time. For each lag k, we regress the fatal violence rate of a market-month on search in the same market k months earlier,

$$\operatorname{asinh}(y_{mt}) = b_k \ln(S_{m,t-k}) + a_m + g_t + e_{mt} \qquad (2)$$

where y_{mt} is fatal interpersonal incidents per 100,000 residents in market m and month t, $S_{m,t-k}$ is the calibrated search index k months earlier, and a_m and g_t are market and month fixed effects. Each market is weighted by its mean population, so the estimator is weighted least squares. With both sets of fixed effects, each market is compared with itself over time, after anything that affects the whole country in a given month has been removed. The estimate therefore comes from differences across markets in when and how much search and violence changed. The lags are constructed on the calendar grid before estimation, so missing months cannot shift the meaning of a lag.

The quantity of interest, however, is the cumulative response over twelve months. We estimate it from a joint distributed-lag model that enters all thirteen lags together:

$$\text{asinh}(y_{mt}) = \sum_{k=0}^{12} b_k \ln(S_{m,t-k}) + a_m + g_t + e_{mt} \qquad (3)$$

The cumulative response is the sum of the lag coefficients in Eq. (3), and we estimate it directly with its own standard error. For the checks in Section 5, we use lag 6 of Eq. (2) as a reference single-lag estimate. Lag 6 lies in the middle of the twelve-month window, and a model with one lag needs search in fewer months, so it keeps more market-months than Eq. (3) (23,426 against 19,950). We estimated it in the same form with every data source and comparison set. Most checks therefore test the lag-6 estimate and not the twelve-month response, and Sections 5.1 and 5.5 also report the twelve-month response without the yearly cycle, with the leads, and in the linear and Poisson models.

Two reporting choices follow from the data. The outcome is transformed with the inverse hyperbolic sine (asinh), a transformation that Burbidge et al. (1988) brought into regression practice. At large values the asinh behaves like a log, and unlike a log it can be applied to months with zero incidents. Its coefficient, however, depends on the units in which the outcome is measured (Chen and Roth 2024), so we use the asinh only as a scale for reporting and also report two models whose tests do not depend on these units, a linear model of the rate and a Poisson count model (Section 5.5). The linear model measures a change in the level of the rate, and the Poisson model a proportional change. Population weights address the unequal error variance that arises when a rate outcome has varying denominators, although they do not by themselves deliver a population average effect when the association varies with market size (Solon et al. 2015). Weighting by population also follows crime-panel practice, in which the estimate is meant to represent where people live (Chalfin and McCrary 2018). In our data the association does vary with market size, so we also report unweighted and size-stratified estimates (Section 5.6; Supplements D.1 and D.5).

Most threats to this design have a check of their own. Since search after the outcome month would predict violence if violence drove later search, we add six leads of search, at t+1 through t+6, to Eq. (3) and test them jointly. Both series also follow a yearly cycle. From 2015 to 2019, search within markets was highest in November and December, and fatal violence was highest in July (Supplement D.11). The month effects remove only the part of this cycle that all markets share. A cycle that repeats every year can line up search with later violence and with earlier violence, so it can produce associations at both lags and leads. We remove the cycle in two ways.

The first uses only the non-seasonal part of search, which is what remains after each market's average for each calendar month of 2015-2019 is subtracted, so it needs search data for those years. The second adds market-by-calendar-month fixed effects, which remove the cycle from both series within the regression. The month effects absorb shocks that hit the whole country at the same time, while shocks that reached some regions earlier than others are addressed in part, with controls for pandemic restrictions, protests, firearm laws, and enforcement and with comparisons narrowed to markets of similar size, in the same census division, or in the same state. A confounder that raised violence of every kind would also raise non-firearm homicide, which we test directly. No check, however, can detect a firearm-specific process that followed the same regional timing as the search wave (Section 6.3).

We cluster standard errors by market, the level at which the regressor is assigned (Abadie et al. 2023), and this also allows for serial correlation within each market (Bertrand et al. 2004). We compute the p-values with a wild cluster bootstrap (WCB) of the t statistic under the null hypothesis, which flips the signs of the residuals cluster by cluster (Cameron et al. 2008). Because markets in the same state share state policies and other state-level shocks, standard errors are also clustered by state (Supplement D.3). For each subperiod with a null estimate, we calculate the minimum detectable effect, that is, the smallest effect that the design could have detected (Bloom 1995). We also estimate Eq. (2) for each lead, and we adjust the p-values of the nineteen single-lead and single-lag estimates in Fig. 2 with the Holm step-down procedure (Holm 1979). We fixed the basket of search terms, the protocol for excluding outcomes, and the calibration procedure before estimation. No analysis was pre-registered, so we treat the analyses of timing and the robustness checks as exploratory.

5. Results

5.1 The cumulative response

Increases in acquisition-related search were followed within a year by increases in fatal gun violence (Table 1; Fig. 2a). In the joint model of Eq. (3), the response over twelve months is +0.101 on the asinh scale (95% CI .021 to .182; t = 2.48). Its wild bootstrap p-value is .026 with market clusters and .028 to .030 with state clusters. We then removed each market's usual seasonal pattern from search (Section 4). The twelve-month cumulative response remains after

this step (+.109, t = 2.59, on the 167 markets with enough pre-2020 data to estimate a seasonal pattern for every calendar month, where the unadjusted response is +.1005; Supplement D.11).

The lag-by-lag estimates describe the timing of the association (Fig. 2b). At lag 0, the association is weak (b = .0133, t = 1.49), and lags 1-3 are individually null. The estimates are larger from lag 4 through lag 9, except for a weaker lag-8 estimate (t = 1.42), and they fall back at lags 10 through 12. Part of this shape comes from the yearly cycles of search and violence (Section 4). With only the non-seasonal part of search, the response spreads out over the year, and about half of it falls in lags 0-3. Single-lag estimates with the cycle removed from both series are smaller, and their intervals include both zero and the baseline (Supplement D.11). The timing of the response within the year is therefore uncertain, and we do not read the lag profile as an estimate of the delay. Of the single-lag coefficients, only lag 9 survives the Holm adjustment (b = .0294, t = 3.37).

For the checks in the rest of this section, we use lag 6 as a reference single-lag estimate, b = .0308 (t = 2.86; market-clustered wild bootstrap p = .0051, G = 210; state-clustered p = .019). This estimate conditions on market-months in which the current search value is also observed, whereas the per-lag profile in Fig. 2 requires only the lagged value and gives .0301 on a slightly larger sample. The lag-6 estimate keeps the yearly cycle in search, so most of the checks below test the association with raw search.

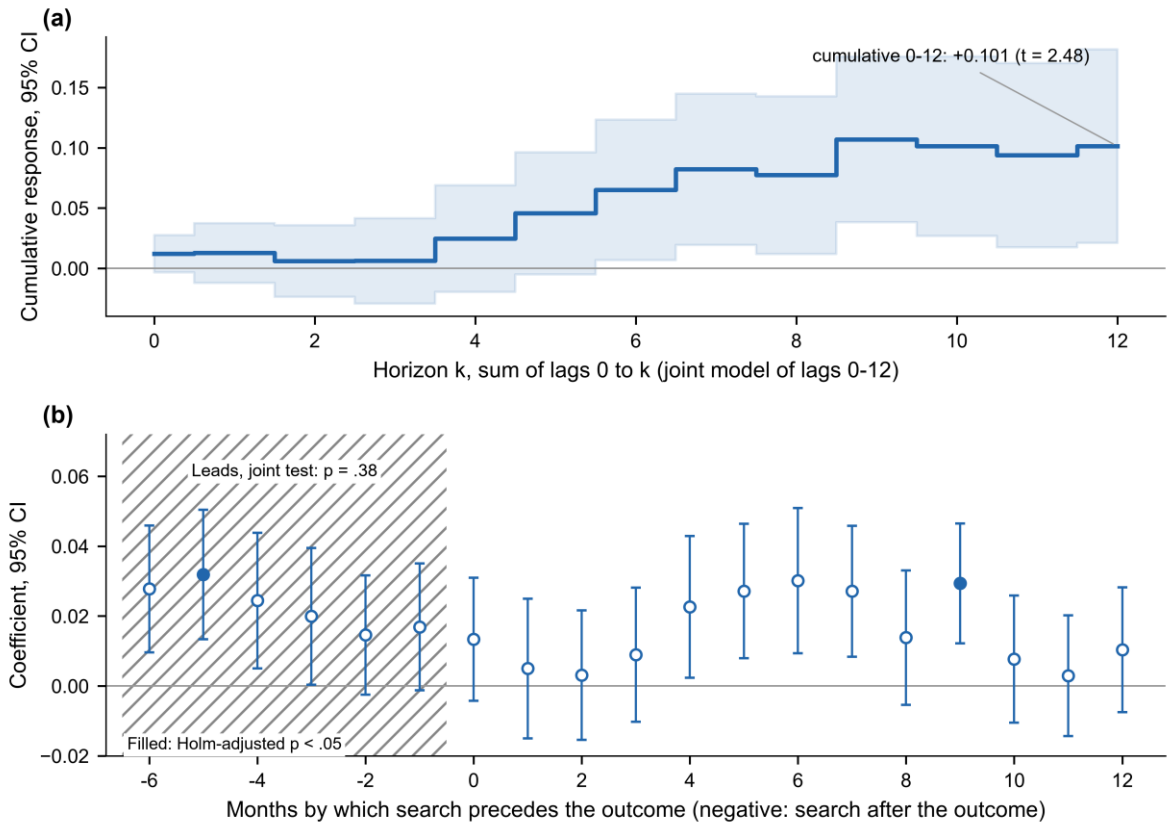


Fig. 2 The association over twelve months. Panel (a) shows the cumulative response from the joint distributed-lag model of Eq. (3), with a 95% market-clustered CI. In panel (b), each point is the coefficient from a separate regression of the outcome at t on ln(search) at t-k, with market and month fixed effects, population weights, and 95% market-clustered CIs. A value of k > 0 means that search precedes the outcome, and the leads are shown as negative k (written t+h in the text) in the hatched region. Filled markers remain significant at the 5 percent level after a Holm adjustment across the 19 estimates in panel (b), applied to wild cluster bootstrap p-values. They are lag 9 and the lead t+5. The p = .38 printed in panel (b) is the joint test of the six leads added to Eq. (3) (Table 1). In that model the six leads sum to +.043 (t = 1.97).

Table 1 Cumulative response and leads of search

Quantity	Result
Panel A. Distributed lag (lags 0-12, joint model)	
Cumulative response, 0-12	+0.101 (.041); 95% CI [.021, .182]; t = 2.48
Panel B. Leads of search	
Leads t+1 to t+6, joint	p = .38
Leads t+1 to t+6, sum	+0.043 (.022); t = 1.97
Leads t+1 to t+6, sum, market-by-calendar-month FE	+0.022 (.023); t = 0.94; joint p = .58
Lead t+5, single coefficient	.0319 (.0094); Holm p = .021

Table notes: Both panels use market and month fixed effects with population weights, and standard errors are clustered by market. Panel A comes from Eq. (3), estimated on the 19,950 market-months, in 192 markets, in which search is observed at all thirteen lags. The 18 markets left out hold 0.65 percent of the population. For the joint lead test and the first six-lead sum, we add the six leads to Eq. (3), so these use only the months in which the leads are also observed (18,556 market-months in 189 markets). The model for the second sum also includes market-by-calendar-month fixed effects, which remove each market's seasonal pattern from both series (18,520 market-months in 184 markets; Supplement D.11). The single-lead coefficient comes from Eq. (2), as in Fig. 2. For joint tests we use F(q, G-1). The joint tests of lags 4-9 and of the remaining lags are in Supplement D.11.

5.2 Three sources of data

Table 2 repeats the reference-lag estimate with each of the three sources of data on fatal firearm violence introduced in Section 3.2. The sign and significance replicate in all three. The levels of the coefficients cannot be compared, because the sources count different events and the vital statistics resolve only to states.

Table 2 The association in three independent measures of firearm violence

Outcome source	Level	Sample	N	b (SE)	t
Gun Violence Archive, fatal incidents (baseline)	market-month	210 markets, 2015-2025	23,426	.0308 (.0108)	2.86
Supplementary Homicide Reports, firearm homicide	market-month	202 markets, 2015-2024 except 2021	18,008	.0324 (.0101)	3.21
Vital statistics, firearm homicide	state-month	state-months with an unsuppressed NVSS count (51 states)	4,338	.0747 (.0279)	2.68

Table notes: Each row uses the lag-6 specification of Eq. (2), with population weights, unit and month fixed effects, and inference clustered on the unit. N is the number of market-months or state-months in the estimation sample. The levels cannot be compared across rows. GVA counts fatal incidents, the homicide reports count homicides for which the weapon is recorded, and the vital statistics count deaths by cause, while the state-month row also uses coarser units than the market-month rows. Supplement B.2 (Table B1) holds the sample fixed and changes only the

source of the outcome. The homicide-report row ends in 2024, the last year that the source covers, and leaves out 2021 and Florida (Section 3.2).

5.3 Where the identifying variation lies

Narrowing the comparison set shows where the identifying variation lies (Table 3, Panel A). Population-decile-by-month effects compare each market only with markets of similar size in the same month and leave regional shocks in the data, and they give b = .0291 (t = 2.57), close to the baseline. Division-by-month effects reduce the estimate to b = .0173 (t = 1.83). State-by-month effects compare each market only with other markets in the same state and month, and their estimate of b = .0126 (SE = .0084) is small and imprecise. Its confidence interval runs from below zero to just under the baseline, and corrected for sampling error in search, the interval also includes the baseline (Supplement D.6). Variation within states therefore neither demonstrates the association nor excludes one that is only slightly smaller than the baseline. Such intervals could come from a true null within states plus a confounder that operates across states, or from an association that is too imprecise to detect at this resolution.

Most of the identifying variation in search comes from differences across markets in when the wave of searches arrived and how it unfolded (Supplement D.6). Because the crests of the wave nearly coincide within each state, these differences run mostly across states and regions. The association is therefore regionally identified, and confounders that arrived earlier or more strongly in some regions would operate on the same margin. The firearm-specificity test of Section 5.4 and the controls of Section 5.5 address such confounders, although only in part. The same model estimated with states in place of media markets gives a larger but less precise coefficient (b = .0531, t = 1.90, wild bootstrap p = .103; Supplement D.7), so the comparison of markets is the more precise of the two.

Table 3 Narrower comparison sets and controls for potential regional confounders

Specification	Comparison set	b (SE)	t
Market + month FE (baseline, lag 6)	all markets, national month shocks removed	.0308 (.0108); Fig. 2b sample .0301 (.0106)	2.86 (WCB p = .0051); 2.85
Panel A. Narrower comparison sets			
Population-decile-by-month FE	similar-size markets, same month	.0291 (.0113)	2.57
Division-by-month FE	markets in the same census division and month	.0173 (.0095)	1.83
State-by-month FE	markets in the same state and month	.0126 (.0084)	1.50
Panel B. Controls for potential regional confounders			
+ pandemic stringency and mobility	same	.0299-.0303	2.85-2.92
+ protest controls (2017-2024 window)	same, 2017-2024 only	.0432 (.0125) vs .0433 uncontrolled	3.46
+ firearm-law-change controls (RAND stocks, six-month lags, new-law flows)	same	.0276-.0304	2.68-2.81
Excluding May 2020-June 2021	same, without these months	.0236 (.0107)	2.20

Table notes: Panel A narrows the comparison set, and Panel B adds controls to the baseline or drops the peak months. Every row except the baseline and the firearm-law row follows the sample rule of the lag-6 estimate in Fig. 2b, which requires search only at t-6 (Section 5.1). The full-sample estimate under this rule is the second value in the first row (N = 24,384). The protest row is estimated on 2017-2024, the window that the protest source covers, and its comparison value is the uncontrolled coefficient on the same window (.0433). The firearm-law row adds the RAND stocks of restrictive and permissive laws, their six-month lags, and the flows of new laws, and the range covers the individual sets of controls. The full control set is at the lower end of that range (Section 5.5).

5.4 Firearm specificity

The association is specific to firearms. A confounder that moved all kinds of violence together would also predict non-firearm homicide, so we contrast firearm with non-firearm homicide (Fig. 3). Over the common window of the vital statistics, from July 2015 through December 2024, the last month of the non-firearm series, the search coefficient is .0784 (t = 2.73) for firearm homicide and .0001 (t = 0.01) for non-firearm homicide. We estimated the difference by stacking the two equations and bootstrapping at the state level, and it is .0783 (t = 2.75; p = .024;

Supplement B.3). At the market level, the homicide reports show the same contrast, .0324 against .0062, a difference of .0263 (bootstrap p = .010; Supplement B.4).

Two qualifications apply to this test. Firstly, the period itself partly produces the contrast. Violence rose in firearm homicide and hardly at all in non-firearm homicide, so an index that tracks the wave of firearm acquisition would fail to predict non-firearm homicide, almost regardless of the channel. Secondly, both versions of the test draw on the same regional variation as the baseline estimate (Section 5.3), so the placebo adds no confirmation within states and cannot separate the acquisition channel from other regional processes specific to firearms (Section 6.3).

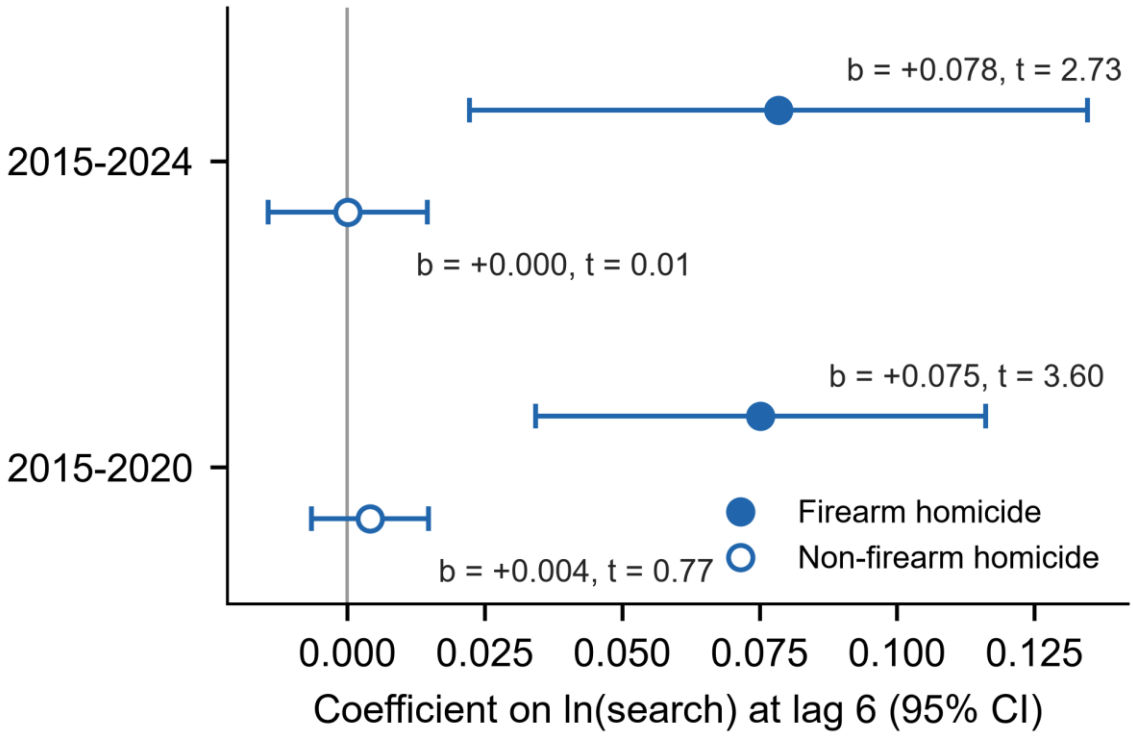


Fig. 3 Firearm specificity. Filled markers show the lag-6 coefficient on ln(search) for firearm homicide, and open markers show it for non-firearm homicide. Both come from the vital-statistics state-month data, the two coefficients in a pair are estimated on the same sample, and the bars are 95% confidence intervals clustered by state. The upper pair uses July 2015 to December 2024, and the lower pair stops in December 2020 (Supplement B.3).

5.5 Robustness

We first ask whether violence came before search. (1) Leads. The six leads of search added to Eq. (3) are not jointly significant ($p = .38$; Table 1). Several single leads are positive, and the lead at t+5 survives the Holm adjustment on its own (Fig. 2b). In the joint model of leads and lags, its coefficient falls to .0172 ($t = 2.04$), from .0319 alone. The yearly cycles of Section 4 line up with leads as well as with lags, so a single lead can pick them up, and the t+5 estimate alone says little about direction. The six leads together sum to +.043 ($t = 1.97$), and the same cycles may account for part of this sum. With market-by-calendar-month fixed effects (Section 4), the six leads sum to +.022 ($t = 0.94$) on nearly the same sample. The twelve-month response is +.074 ($t = 1.89$) in the model with the raw leads and +.088 ($t = 2.32$) with the fixed effects (Supplement D.11). The leads therefore do not show that violence came first, but they do not rule it out.

Controls for potential regional confounders change the lag-6 estimate little, and dropping the peak months lowers it without removing it (Table 3, Panel B). (2) Pandemic conditions. With state pandemic restrictions and Google mobility in the model, the estimate is .0299 to .0303. (3) Protests. On 2017-2024, the window that the protest data cover, the estimate is .0432 with protest controls and .0433 without them. (4) Firearm laws. Adding the RAND stocks of restrictive and permissive laws, their six-month lags, and the flows of new laws gives .0276 to .0304, and the full set of law controls gives .0276 ($t = 2.68$). (5) Enforcement. With counts of officer-involved shootings and state-month arrests for weapons, drug, and all offenses in the model, the per-lag estimate moves from .0301 to between .0276 and .0303 (Supplement D.2). (6) The peak months. Excluding May 2020 through June 2021 lowers the estimate to .0236 ($t = 2.20$). The law and enforcement variables are state-level quantities or equilibrium outcomes, and the adoption of laws may itself respond to violence, so these results show how far conditioning on them moves the estimate, but they do not decompose the association. Within those limits, changes in firearm law do not account for the association.

The last check concerns the specification of the model. (7) Units and counts. A linear model of the untransformed rate, whose t statistic does not depend on the units of the outcome, gives $t = 2.82$ at lag 6 and $t = 2.41$ for the cumulative response. Because a Poisson count model already gives larger markets more weight through its population offset, we estimate it without further weights, and its lag-6 and cumulative estimates have the same sign as the baseline but are not significant ($t = 1.80$ and 1.50; Supplement D.4).

5.6 Boundaries in time and market size

The association is bounded in time and in market size (Table 4). We look first at time. For 2015-2019, the coefficient is null (b = .0023, t = 0.24). Search still tracked background checks in those years, but the slope of that link was smaller (Section 3.1), so part of this null may come from the weaker link between search and purchases. The 95 percent confidence interval of the 2015-2019 coefficient is [-.016, .021], so it excludes the pooled estimate. The pre-2020 years are therefore inconsistent with the pooled effect, although a smaller effect cannot be ruled out (Supplement D.4). Power was sufficient for this question, as the minimum detectable effect of the subperiod at 80 percent power is .0264, which is below the pooled estimate. The estimates for 2020-2021 (b = .0654, t = 3.88) and 2022-2025 (b = .0285, t = 2.33), on the other hand, are both positive and significant. When search is split into its seasonal and non-seasonal parts, the non-seasonal part gives a surge estimate of b = .0519 (t = 2.92) and a smaller, non-significant estimate for 2022-2025 (b = .0199, t = 1.62; Supplement D.11). The evidence for 2022-2025 therefore depends on whether the seasonal part of search is kept, while the surge estimate holds either way.

Regarding market size, the association is strongest in the largest markets. Among the twenty-five largest markets, it appears even without population weights (b = .066, t = 3.10). In the full unweighted sample it does not appear (t = 1.09), and it fades below roughly the fiftieth market (Supplement D.5). The fifty largest markets are home to about two-thirds of the US population (68 percent).

Table 4 Boundaries of the association

Subsample	**b (SE)**	**t**
Panel A. Time		
2015-2019	.0023 (.0094)	0.24
2020-2021	.0654 (.0169)	3.88
2022-2025	.0285 (.0122)	2.33
Panel B. Market size		
Top-25 markets, unweighted	.066 (.0213)	3.10
Full sample, unweighted	.0098 (.0090)	1.09

Table notes: Both panels use the lag-6 specification of Eq. (2) under the sample rule of the lag-6 estimate in Fig. 2b, which requires search only at t-6 (Section 5.1). The weighted full-sample estimate under this rule is .0301 (N = 24,384). Equivalence tests for 2015-2019 are in Supplement D.4.

6. Discussion

6.1 What the association means

The association is consistent with some newly bought guns reaching violence within a year. Because the estimate is regionally identified, a firearm-specific process with the same regional timing would produce the same pattern (Sections 5.3, 5.4, and 6.3). A causal reading requires that some of the new acquisitions reach misuse, either by the buyers themselves or after a transfer. The trace data allow both routes. Most guns recovered within a year of sale are recovered in the state where they were sold (Section 2.1). The consequences of a purchase wave therefore remain largely within the regions that the design compares.

Where the delay falls matters for this reading. Brownstein et al. (2020) examined lags of up to six months in either direction in national monthly data. They found that search volume and background checks were most highly correlated at zero lag. Our basket of seven terms also peaks at zero lag, with a clean single peak in the within-state correlation with background checks, and the peak stays at zero lag when month-of-year means are removed from both series (Supplement A.7). Background checks exist only by state and month, so we assume that the zero-lag result also holds within markets, where no series of checks can verify it. If it does, the delay falls between purchase and violence, and not between search and purchase. Whether search leads violence does not depend on this assumption, but reading the delay as an interval from transaction to violence does. A rise in violence that begins within a year of a rise in search fits better with more direct routes, such as misuse by the purchasers themselves or diversion close to retail, than with the slower routes of most traced crime guns (Section 2.1).

6.2 Relation to earlier studies

In Levine and McKnight (2017), homicide did not respond to the spike in gun sales (Section 2.1), but their result can be reconciled with ours. Their window covered the five months after the Sandy Hook shooting, and they noted that their design identifies short-run effects. Other purchase spikes studied before 2020 were also short, and their estimated effects on homicide were mixed (Section 2.1). Our design detects no association in the years before 2020 either (Section 5.6). The association in our data comes mainly from the 2020-2021 surge, which lasted almost two years.

The state-level test of Schleimer et al. (2021), whose window closed in mid-2020 (Section 2.1), is also consistent with our estimates. With states in place of media markets, our lag-6 coefficient is only marginal (Section 5.3), so our market-level association does not overturn their state-level null for non-domestic firearm violence. They did report a positive result for firearm injuries from domestic violence. Excess purchasing was associated with these injuries in April and May 2020 (rate ratios of 2.60 and 1.79 per excess firearm purchase per 100 population), but the estimates weakened and their intervals crossed the null when state-specific linear trends were added. Fazel-Zarandi and Barnett (2024) compared homicide growth across 50 large cities in a single year, 2020. Our estimates follow each market month by month from 2015 to 2025, so the two designs draw on different variation.

Researchers who doubt that gun ownership raises crime have set tests for studies of this kind, and two of them apply to our measure. Kovandzic et al. (2013) argued that a stand-in for gun ownership is valid only if it moves over time with a direct measure of ownership (Section 2.1). We can apply this test to the search index, with background checks as the direct measure. Within states, the index and background checks have a median correlation of .78 (Section 3.1), although background checks are themselves a stand-in for purchases (Section 6.3). Kleck (2015) argued that using an earlier value of a predictor does not show which came first (Section 2.1). Our evidence on order does not rest on the lag alone, because search in the six months that follow, added to the model, is not jointly significant, although this test is not decisive (Section 5.5).

Before 2020, we detect no association between search and violence, although search was a weaker guide to purchases in those years (Section 5.6). Kleck (2015) tentatively concluded, for the stock of guns, that higher gun ownership rates do not cause higher crime rates, and our null in these years resembles that conclusion. The association should therefore not be read as a constant relationship that also holds in years without a purchase surge.

One tension remains with the protective estimate of Kovandzic et al. (2013), which they attributed mainly to guns held by noncriminals (Section 2.1). Their parameter describes the standing stock of guns, whereas the association here describes a shock to the flow of guns. Because the two studies estimate different quantities, their findings do not contradict each other directly.

6.3 Limitations

This study has several limitations. Firstly, search is not a purchase record, and the basket also includes "gun safe" and "gun maintenance", so the share of the index that comes from these storage and maintenance phrases may differ across markets and shift over time, and that share cannot be seen in current search volumes. Secondly, background checks are themselves a proxy for purchases (Brownstein et al. 2020), and the validation can be no better than this criterion. We had no purchase record below the state level to serve as a criterion for the market index. We therefore checked the market index for reliability and for agreement with the state panel and with the 2022 retrieval, and we compared it with background checks after summing the markets to states (Section 3.1). Third, GVA counts come from media and police reports, and in an event-level audit of four cities, Gobaud et al. (2023) found good but incomplete capture with systematic missingness, although the close correlation with vital statistics in Section 3.2 limits this concern for fatal counts. Another limitation is that the evidence comes from a single platform, whose updates can change what the same query returns over time. An update that raises or lowers every market by the same proportion in a month is absorbed by the month fixed effects, in the same way as the monthly scale (Section 3.1). Only an update that affects markets differently, for example by moving small markets across the reporting threshold, could bias the estimate. A retrieval of the same phrases in early 2022, made with different software, correlates with the present panel at the state and market levels reported in Section 3.1 (Supplement A.7). Used in place of the present panel for 2015-2021, it also gives a similar estimate, b = .0202 against .0240 for the present panel on the same market-months.

The design is also ecological (Morgenstern 1995). We observe search and violence only as totals for each market and month, so we cannot tell whether the people who searched for or bought guns were the people involved in the violence. Our units are also large, because gun violence concentrates in a small share of street segments and intersections (Braga et al. 2010), and a media market covers a whole metropolitan region. The evidence also comes from one country and one purchase surge.

The fixed effects remove stable differences between markets and changes that reach all markets in the same month. They do not remove changes that reached regions at different times, and our estimate rests largely on such regional differences (Section 5.3). Regional confounding therefore cannot be ruled out. Enforcement, courts, drug markets, illegal carrying, and the capacity for

community violence intervention may all have changed on regional schedules, and no direct market-month measure of these changes exists. The firearm-specificity test cannot rule out such a confounder. A change that raised only firearm violence and reached each region at the time of its purchasing wave would pass the test (Section 5.4), and we regard such a change as the main threat to our identification. We can say that violence rose within a year of a rise in search, but not in which month of that year the rise was largest. Part of the variation we use comes from the yearly cycles of search and violence, and once these cycles are removed, the response spreads across the year (Section 5.1).

Finally, our standard errors treat markets as independent of one another. In Chicago, shooting trends move together even in neighborhoods that do not border each other (Aleinzi et al. 2025), so markets may not be fully independent either. Standard errors clustered by state, which allow markets in the same state to be correlated, lead to the same conclusion (Supplement D.3). Spatial dependence is not fully modeled here, but the test in Supplement D.3 finds that the dependence left after the month effects is detectable but small.

6.4 What the measure is for

The value of the measure comes from the gaps that federal law leaves in the record of gun purchases (Section 1). Where purchase records exist, as in the dealer systems of single states (Wright et al. 2010), they are better than any proxy, and if comprehensive records became available at the national level, the panel would lose most of its purpose.

The estimated association can also be checked against actual transactions. The point-of-sale system of at least one large state, such as the California system used by Wright et al. (2010), could test it against transaction records. Transaction records would also show at which lags within the year the association appears.

7. Conclusion

In conclusion, acquisition-related search can be measured for each media market and month, and within states it tracks background checks (Section 3.1). Across large US media markets from 2015 to 2025, fatal gun violence rose within a year after increases in this search, whereas non-firearm homicide did not, and the association was strongest during the 2020-2021 surge and was not detected before 2020. As the association rests on regional differences in the timing of the

search wave, a firearm-specific process with the same regional timing remains the main alternative explanation. Purchase records from a state point-of-sale system could test, for that state's markets, whether the association reflects actual purchases.

Declarations

Funding. The authors received no specific funding for this work.

Competing interests. The authors have no financial or non-financial interests directly or indirectly related to this work to disclose.

Data availability. The calibrated search panels, the pass-level raw retrievals behind them, all analysis code, and aggregated outcome panels will be deposited in a public repository with a DOI upon acceptance. Incident-level Gun Violence Archive data were provided by the Gun Violence Archive and cannot be redistributed under the data provider's terms; aggregated market-month panels sufficient to reproduce every estimate are included in the replication package, consistent with published guidance on aggregation-based sharing for sensitive crime data (Campedelli et al. 2026). Vital-statistics benchmarks were exported from CDC WONDER. Query specifications are included in the package.

Ethics approval. The study uses only aggregated secondary data on incidents compiled from public sources and platform-level search indices; no human participants were recruited, and no individual-level data were analyzed.

AI Use Disclosure. We used Anthropic's Claude (Fable 5.1 & Opus 5.5) to assist with coding, specifically to retrieve the Google Trends data, build the analysis panels (adapting our existing code), and run the statistical analyses and robustness checks, as well as for grammar checking. The authors directed this work, reviewed all output, and take full responsibility for the content of the article.

References

Abadie A, Athey S, Imbens GW, Wooldridge JM (2023) When should you adjust standard errors for clustering? Q J Econ 138:1-35. https://doi.org/10.1093/qje/qjac038

Aleinzi M, Sharkey P, Pryce G, Hughes N (2025) Spatial networks of neighborhood violence. J Quant Criminol. https://doi.org/10.1007/s10940-025-09637-3

Anderberg D, Rainer H, Siuda F (2022) Quantifying domestic violence in times of crisis: an internet search activity-based measure for the COVID-19 pandemic. J R Stat Soc Ser A Stat Soc 185:498-518. https://doi.org/10.1111/rssa.12780

Askitas N, Zimmermann KF (2009) Google econometrics and unemployment forecasting. Appl Econ Q 55:107-120. https://doi.org/10.3790/aeq.55.2.107

Azrael D, Hepburn L, Hemenway D, Miller M (2017) The stock and flow of U.S. firearms: results from the 2015 National Firearms Survey. RSF Russell Sage Found J Soc Sci 3(5):38-57. https://doi.org/10.7758/RSF.2017.3.5.02

Berniell I, Facchini G (2021) COVID-19 lockdown and domestic violence: evidence from internet-search behavior in 11 countries. Eur Econ Rev 136:103775. https://doi.org/10.1016/j.euroecorev.2021.103775

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? Q J Econ 119:249-275. https://doi.org/10.1162/003355304772839588

Bestenbostel A, Peralta A (2026) WARNings - the impact of negative economic news on racial animus. J Econ Behav Organ 243:107399. https://doi.org/10.1016/j.jebo.2025.107399

Bleher J, Dimpfl T (2022) Knitting multi-annual high-frequency Google Trends to predict inflation and consumption. Econom Stat 24:1-26. https://doi.org/10.1016/j.ecosta.2021.10.006

Bloom HS (1995) Minimum detectable effects: a simple way to report the statistical power of experimental designs. Eval Rev 19:547-556. https://doi.org/10.1177/0193841X9501900504

Braga AA, Papachristos AV, Hureau DM (2010) The concentration and stability of gun violence at micro places in Boston, 1980-2008. J Quant Criminol 26:33-53. https://doi.org/10.1007/s10940-009-9082-x

Braga AA, Wintemute GJ, Pierce GL, Cook PJ, Ridgeway G (2012) Interpreting the empirical evidence on illegal gun market dynamics. J Urban Health 89:779-793. https://doi.org/10.1007/s11524-012-9681-y

Brownstein JS, Nahari AD, Reis BY (2020) Internet search patterns reveal firearm sales, policies, and deaths. npj Digit Med 3:152. https://doi.org/10.1038/s41746-020-00356-6

Burbidge JB, Magee L, Robb AL (1988) Alternative transformations to handle extreme values of the dependent variable. J Am Stat Assoc 83:123-127. https://doi.org/10.1080/01621459.1988.10478575

Cameron AC, Gelbach JB, Miller DL (2008) Bootstrap-based improvements for inference with clustered errors. Rev Econ Stat 90:414-427. https://doi.org/10.1162/rest.90.3.414

Campedelli GM, Jelveh Z, Chalfin A, Semenza D, Piza E, Albors Zumel A, Lepri B, Sharkey P (2026) Computational approaches and the future of urban crime research. Nature 655:315-326. https://doi.org/10.1038/s41586-026-10622-4

Caputi TL, Ayers JW, Dredze M, Suplina N, Burd-Sharps S (2020) Collateral crises of gun preparation and the COVID-19 pandemic: infodemiology study. JMIR Public Health Surveill 6:e19369. https://doi.org/10.2196/19369

Centers for Disease Control and Prevention (2021) National Vital Statistics System, Mortality 1999-2020 on CDC WONDER Online Database, released in 2021. National Center for Health Statistics. Data are from the Multiple Cause of Death Files, 1999-2020, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. http://wonder.cdc.gov/ucd-icd10.html. Accessed 16 July 2026

Centers for Disease Control and Prevention (2026a) National Vital Statistics System, Mortality 2018-2024 on CDC WONDER Online Database, released in 2026. National Center for Health Statistics. Data are from the Multiple Cause of Death Files, 2018-2024, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. http://wonder.cdc.gov/ucd-icd10-expanded.html. Accessed 16 July 2026

Centers for Disease Control and Prevention (2026b) National Vital Statistics System, Provisional Mortality on CDC WONDER Online Database. National Center for Health Statistics. Data are from the final Multiple Cause of Death Files, 2018-2024, and from provisional data for years 2025 and later, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. http://wonder.cdc.gov/mcd-icd10-provisional.html. Accessed 16 July 2026

Chalfin A, McCrary J (2018) Are U.S. cities underpoliced? Theory and evidence. Rev Econ Stat 100:167-186. https://doi.org/10.1162/rest_a_00694

Chen J, Roth J (2024) Logs with zeros? Some problems and solutions. Q J Econ 139:891-936. https://doi.org/10.1093/qje/qjad054

Chenoweth E, Hammam S, Pressman J, Ulfelder J (2025a) Crowd Counting Consortium U.S. protest event data, 2021-2024, V2. Harvard Dataverse. https://doi.org/10.7910/DVN/9MMYDI

Chenoweth E, Pressman J, Ulfelder J (2025b) Crowd Counting Consortium U.S. protest event data, 2017-2020, V1. Harvard Dataverse. https://doi.org/10.7910/DVN/6OPP7H

Choi H, Varian H (2012) Predicting the present with Google Trends. Econ Rec 88(s1):2-9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Consolidated Appropriations Act, 2010, Pub L No 111-117, div B, tit II, 123 Stat 3034, 3128-29 (2009) (Tiahrt Amendment; reprinted in note to 18 USC § 923)

Cook PJ (2018) Gun markets. Annu Rev Criminol 1:359-377. https://doi.org/10.1146/annurev-criminol-032317-092149

Cook PJ, Ludwig J (2006) The social costs of gun ownership. J Public Econ 90:379-391. https://doi.org/10.1016/j.jpubeco.2005.02.003

Cook PJ, Ludwig J (2019) The social costs of gun ownership: a reply to Hayo, Neumeier, and Westphal. Empir Econ 56:13-22. https://doi.org/10.1007/s00181-018-1497-5

Cook PJ, Ludwig J, Venkatesh S, Braga AA (2007) Underground gun markets. Econ J 117:F588-F618. https://doi.org/10.1111/j.1468-0297.2007.02098.x

Cook PJ, Pollack HA, White K (2019) The last link: from gun acquisition to criminal use. J Urban Health 96:784-791. https://doi.org/10.1007/s11524-019-00358-0

Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297-334. https://doi.org/10.1007/BF02310555

Degli Esposti M, Schell TL, Smart R (2025) The recent rise in homicide: an analysis of weekly mortality data, United States, 2018-2022. Epidemiology 36:174-182. https://doi.org/10.1097/EDE.0000000000001815

Duggan M (2001) More guns, more crime. J Polit Econ 109:1086-1114. https://doi.org/10.1086/322833

Edwards G, Nesson E, Robinson JJ, Vars F (2018) Looking down the barrel of a loaded gun: the effect of mandatory handgun purchase delays on homicide and suicide. Econ J 128:3117-3140. https://doi.org/10.1111/ecoj.12567

Eichenauer VZ, Indergand R, Martínez IZ, Sax C (2022) Obtaining consistent time series from Google Trends. Econ Inq 60(2):694-705. https://doi.org/10.1111/ecin.13049

Fazel-Zarandi MM, Barnett A (2024) Why did US urban homicide spike in 2020? A cross-sectional data analysis for the largest American cities. Risk Anal 44:1616-1629. https://doi.org/10.1111/risa.14271

Federal Bureau of Investigation (2021) UCR Program Quarterly 21-1, February 2021. https://le.fbi.gov/file-repository/ucr-program-quarterly-february-2021.pdf. Accessed 25 October 2026

Federal Bureau of Investigation (2026) Crime Data Explorer: arrest counts by state and month, Uniform Crime Reporting Program, retrieved through the Crime Data Explorer API. https://cde.ucr.cjis.gov/. Accessed 20 July 2026

Firearms Owners' Protection Act, Pub L No 99-308, § 106, 100 Stat 449, 459-460 (1986) (codified at 18 USC § 926(a))

Franzén A (2023) Big data, big problems: why scientists should refrain from using Google Trends. Acta Sociol 66:343-347. https://doi.org/10.1177/00016993221151118

Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) Detecting influenza epidemics using search engine query data. Nature 457:1012-1014. https://doi.org/10.1038/nature07634

Gobaud AN, Mehranbod CA, Kaufman E, Jay J, Beard JH, Jacoby SF, Branas CC, Bushover B, Morrison CN (2023) Assessing the Gun Violence Archive as an epidemiologic data source for community firearm violence in 4 US cities. JAMA Netw Open 6(6):e2316545. https://doi.org/10.1001/jamanetworkopen.2023.16545

Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ (2010) Predicting consumer behavior with Web search. Proc Natl Acad Sci USA 107:17486-17490. https://doi.org/10.1073/pnas.1005962107

Google LLC (2022) Google COVID-19 community mobility reports. https://www.google.com/covid19/mobility/

Gun Violence Archive (2026a) Incident-level data, 1 January 2015 to 31 December 2025, bulk extract supplied to the authors by the Gun Violence Archive, July 2026. https://www.gunviolencearchive.org

Gun Violence Archive (2026b) General methodology. https://www.gunviolencearchive.org/methodology. Accessed 17 August 2026

Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, Webster S, Cameron-Blake E, Hallas L, Majumdar S, Tatlow H (2021) A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 5(4):529-538. https://doi.org/10.1038/s41562-021-01079-8

Hayo B, Neumeier F, Westphal C (2019) The social costs of gun ownership revisited. Empir Econ 56:1-12. https://doi.org/10.1007/s00181-018-1496-6

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65-70. https://www.jstor.org/stable/4615733

Hölzl J, Keusch F, Sajons C (2025) The (mis)use of Google Trends data in the social sciences - a systematic review, critique, and recommendations. Soc Sci Res 126:103099. https://doi.org/10.1016/j.ssresearch.2024.103099

Johnson D, Elliott S, Dooley BD, Stickle BF (2026) Is gun demand national? Evidence from Google searches and NICS background checks. J Quant Criminol. https://doi.org/10.1007/s10940-026-09679-1

Kaplan J (2025) Jacob Kaplan's concatenated files: Uniform Crime Reporting Program data: Supplementary Homicide Reports, 1976-2024. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E100699V16

Kegler SR, Simon TR, Zwald ML, Chen MS, Mercy JA, Jones CM, Mercado-Crespo MC, Blair JM, Stone DM, Ottley PG, Dills J (2022) Vital Signs: changes in firearm homicide and suicide rates - United States, 2019-2020. MMWR Morb Mortal Wkly Rep 71:656-663. https://doi.org/10.15585/mmwr.mm7119e1

Kellermann AL, Rivara FP, Rushforth NB, Banton JG, Reay DT, Francisco JT, Locci AB, Prodzinski J, Hackman BB, Somes G (1993) Gun ownership as a risk factor for homicide in the home. N Engl J Med 329:1084-1091. https://doi.org/10.1056/NEJM199310073291506

Kleck G (2015) The impact of gun ownership rates on crime rates: a methodological review of the evidence. J Crim Justice 43:40-48. https://doi.org/10.1016/j.jcrimjus.2014.12.002

Koenig C, Schindler D (2023) Impulse purchases, gun ownership, and homicides: evidence from a firearm demand shock. Rev Econ Stat 105:1271-1286. https://doi.org/10.1162/rest_a_01106

Kovandzic T, Schaffer ME, Kleck G (2013) Estimating the causal effect of gun prevalence on homicide rates: a local average treatment effect approach. J Quant Criminol 29:477-541. https://doi.org/10.1007/s10940-012-9185-7

Laqueur HS, Kagawa RMC, McCort CD, Pallin R, Wintemute G (2019) The impact of spikes in handgun acquisitions on firearm-related harms. Inj Epidemiol 6:35. https://doi.org/10.1186/s40621-019-0212-0

Laqueur HS, McCort C, Smirniotis C, Robinson S, Wintemute GJ (2023) Trends and sources of crime guns in California: 2010-2021. J Urban Health 100:879-891. https://doi.org/10.1007/s11524-023-00741-y

Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google Flu: traps in big data analysis. Science 343:1203-1205. https://doi.org/10.1126/science.1248506

Levine PB, McKnight R (2017) Firearms and accidental deaths: evidence from the aftermath of the Sandy Hook school shooting. Science 358:1324-1328. https://doi.org/10.1126/science.aan8179

Liu K (2024) The measurement errors of google trends data. Discov Data 2:7. https://doi.org/10.1007/s44248-024-00013-3

Liu Y-H, Wolff KT, Lo T-Y (2023) Big data in crime statistics: using Google Trends to measure victimization in designated market areas across the United States. Methodol Innov 16(3):341-358. https://doi.org/10.1177/20597991231183962

Lu R, Sheng SY (2022) How racial animus forms and spreads: evidence from the coronavirus pandemic. J Econ Behav Organ 200:82-98. https://doi.org/10.1016/j.jebo.2022.05.014

Luca M, Malhotra D, Poliquin C (2017) Handgun waiting periods reduce gun deaths. Proc Natl Acad Sci USA 114:12162-12165. https://doi.org/10.1073/pnas.1619896114

Mellon J (2014) Internet search data and issue salience: the properties of Google Trends as a measure of issue salience. J Elect Public Opin Parties 24(1):45-72. https://doi.org/10.1080/17457289.2013.846346

Miller M, Zhang W, Azrael D (2022) Firearm purchasing during the COVID-19 pandemic: results from the 2021 National Firearms Survey. Ann Intern Med 175:219-225. https://doi.org/10.7326/M21-3423

Morgenstern H (1995) Ecologic studies in epidemiology: concepts, principles, and methods. Annu Rev Public Health 16:61-81. https://doi.org/10.1146/annurev.pu.16.050195.000425

Morral AR, Agniel D, Smart R, Schell TL (2024) Household firearm ownership and firearm mortality. JAMA Netw Open 7:e2429335. https://doi.org/10.1001/jamanetworkopen.2024.29335

National Research Council, Institute of Medicine (2013) Priorities for research to reduce the threat of firearm-related violence. National Academies Press, Washington, DC. https://doi.org/10.17226/18319

Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L (2013) Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol 9:e1003256. https://doi.org/10.1371/journal.pcbi.1003256

RAND Corporation (2026) RAND state firearm law database, version 6.0. RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/tools/TLA243-2-v4.html

Raubenheimer J (2024) Of babies, bathwater, and big data: going beneath the surface of Franzén's (2023) Google Trends recommendations. Acta Sociol 67:251-256. https://doi.org/10.1177/00016993231187489

Schleimer JP, McCort CD, Shev AB, Pear VA, Tomsich E, De Biasi A, Buggs S, Laqueur HS, Wintemute GJ (2021) Firearm purchasing and firearm violence during the coronavirus pandemic in the United States: a cross-sectional study. Inj Epidemiol 8:43. https://doi.org/10.1186/s40621-021-00339-5

Sharkey P, Cristancho JC, Semenza D (2024) Deer hunting season and firearm violence in US rural counties. JAMA Netw Open 7:e2427683. https://doi.org/10.1001/jamanetworkopen.2024.27683

Siegel M, Ross CS, King C (2013) The relationship between gun ownership and firearm homicide rates in the United States, 1981-2010. Am J Public Health 103:2098-2105. https://doi.org/10.2105/AJPH.2013.301409

Solon G, Haider SJ, Wooldridge JM (2015) What are we weighting for? J Hum Resour 50:301-316. https://doi.org/10.3368/jhr.50.2.301

Sood G (2016) Geographic information on designated media markets. Harvard Dataverse. https://doi.org/10.7910/DVN/IVXEHT. Accessed 26 January 2020

Stephens-Davidowitz S (2014) The cost of racial animus on a black candidate: evidence using Google search data. J Public Econ 118:26-40. https://doi.org/10.1016/j.jpubeco.2014.04.010

Studdert DM, Zhang Y, Rodden JA, Hyndman RJ, Wintemute GJ (2017) Handgun acquisitions in California after two mass shootings. Ann Intern Med 166:698-706. https://doi.org/10.7326/M16-1574

Studdert DM, Zhang Y, Holsinger EE, Prince L, Holsinger AF, Rodden JA, Wintemute GJ, Miller M (2022) Homicide deaths among adult cohabitants of handgun owners in California, 2004 to 2016: a cohort study. Ann Intern Med 175:804-811. https://doi.org/10.7326/M21-3762

US Bureau of Labor Statistics (2026) Local Area Unemployment Statistics: county monthly data, series file la.data.64.County. https://download.bls.gov/pub/time.series/la/. Accessed 14 July 2026

US Census Bureau (2020) Annual resident population estimates, estimated components of resident population change, and rates of the components of resident population change for states and counties: April 1, 2010 to July 1, 2019 (CO-EST2019-ALLDATA). Population Division, US Census Bureau, Washington, DC

US Census Bureau (2026) Annual resident population estimates, estimated components of resident population change, and rates of the components of resident population change for states and counties: April 1, 2020 to July 1, 2025 (CO-EST2025-ALLDATA). Population Division, US Census Bureau, Washington, DC

US Department of Justice (2023) Justice Department announces publication of second volume of National Firearms Commerce and Trafficking Assessment. Press release 23-128, Office of Public Affairs, 1 February 2023. https://www.justice.gov/archives/opa/pr/justice-department-announces-publication-second-volume-national-firearms-commerce-and. Accessed 25 October 2026

Vose RS, Applequist S, Squires M, Durre I, Menne MJ, Williams CN Jr, Fenimore C, Gleason K, Arndt D (2014) NOAA monthly U.S. climate divisional database (NClimDiv), county temperature and precipitation files. NOAA National Climatic Data Center. https://doi.org/10.7289/V5M32STR. Accessed 14 July 2026

Webster D, Crifasi CK, Vernick JS (2014) Effects of the repeal of Missouri's handgun purchaser licensing law on homicides. J Urban Health 91:293-302. https://doi.org/10.1007/s11524-014-9865-8

Weigend Vargas E, Goldstick JE (2026) Geographic patterns of firearms with short time to crime in the U.S. and the Americas, 2015-2023. Am J Prev Med 70:108310. https://doi.org/10.1016/j.amepre.2026.108310

West R (2020) Calibration of Google Trends time series. In: Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM). Association for Computing Machinery, pp 2257-2260. https://doi.org/10.1145/3340531.3412075

Wright MA, Wintemute GJ, Webster DW (2010) Factors affecting a recently purchased handgun's risk for use in crime under circumstances that suggest gun trafficking. J Urban Health 87:352-364. https://doi.org/10.1007/s11524-010-9437-5

Supplementary Materials

Firearm Acquisition-Related Online Search and Gun Violence across the COVID Purchasing Cycle, 2015-2025

Appendix A. Construction and validation of the search measure

This appendix documents how we built the search measure and how we checked it. Table A1 summarizes the validation of the search measure (Panel A) and of the violence outcome (Panel B). Sections A.1 to A.9 give the detail behind Panel A, and Appendix B gives the detail behind Panel B.

In this supplement, search refers to the Google Trends index for a basket of seven phrases ("how buy gun", "where buy gun", "gun sales", "gun safe", "gun maintenance", "gun shop near me", and "gun store near me"). We measure it in each of the 210 US media markets, which are the designated market areas defined by Nielsen, and some checks use the 51 states instead. The violence outcome comes from the Gun Violence Archive (GVA), and the two comparison sources are the National Vital Statistics System (NVSS) and the FBI's Supplementary Homicide Reports (SHR). Background checks come from the National Instant Criminal Background Check System (NICS). The baseline specification is the lag-6 model of main-text Eq. (2), estimated on the 23,426 market-months in which search is observed both at t and at t-6 (b = .0308). Several checks use the sample rule of the lag-6 estimate in main-text Fig. 2b instead, which requires search only at t-6 (b = .0301, N = 24,384), so a baseline of .0301 below always refers to this rule. We call 2015-2019 the pre-surge years, 2020-2021 the surge, and 2022-2025 the retreat.

Table A1 Validation of the search measure and the violence outcome

Check	What is compared	Level / N	Result
Panel A. Search measure			
Reliability across sampling passes	the thirty retrievals of the same panel, Cronbach alpha with passes as items	17,162 market-months, 6,532 state-months observed in every pass	.97 market; .997 state
Criterion, time series	state search series against NICS background checks, clean definition	50 states, monthly	median within-state r = .78
Criterion, cross-section	search against NICS checks per capita across states, one cross-section per month	monthly cross-sections	mean Spearman rho = .65
Aggregation consistency	market panel aggregated to states with population weights, against the state panel retrieved directly from the platform	48 states (each market assigned to the state holding the largest share of its population), monthly	median within-state r = .98
Independent 2022 retrieval	independent 2022 scrape against the present panel, one cross-section per month	84-month overlap window	median r = .887 across states; .639 across markets
Calibration, blind test	each anchor held out in turn, calibrated from the other four, against its own directly retrieved full-period series	5 market anchors, 5 state anchors	r = .997-.9996 markets; .979-.995 states
Calibration invariance	main coefficient estimated on the calibrated and on the uncalibrated panel	main specification	identical to six decimals
Panel B. Violence outcome			
Criterion, time series	GVA firearm deaths against NVSS firearm homicide (X93-X95)	3,904 state-months, 32 evaluable states	pooled r = .98; within-state r = .93; median within-state r = .908
Cross-source, homicide reports	GVA fatal-incident rate against SHR firearm-homicide rate	19,968 market-months, 202 markets, 2015-2024 except 2021	pooled r = .72; within-market r = .66; median within-market r = .75

Table notes: Pooled correlations use every unit-month at once. Within-unit correlations first remove each unit's mean, so that only movement over time within each unit remains.

A.1 Retrieval pipeline

Each value in the search panel is the average of up to thirty separate retrievals of the same monthly table. Values in a monthly table can be compared across markets within a month, but each month has its own arbitrary scale, and each retrieval is drawn from a sample of searches. We therefore retrieved each of the 132 monthly tables thirty times, in random order, by replaying in a real browser the requests that the platform's own page sends.

One concern with repeated retrievals is that the platform may return the same sample to requests made close together in time. To avoid acquiring the same sample within a short time, we paused

at least four hours between consecutive passes, so consecutive retrievals of the same table were a median of 7.7 hours apart. Identical results are rare in our data. Only 0.3 percent of the 57,420 pairs of retrievals of the same monthly table are identical, and nearly all of these were retrieved less than 12 hours apart. We keep every pass in the average, and a market reported in only some passes takes the mean of the passes that report it. Leaving the identical pairs aside, two retrievals of the same table correlate at r = .866 on average for markets and at .956 for states. After we remove each retrieval's own scale, the sampling errors correlate only between adjacent passes, at +.026. Passes further apart sit near -.034, the value that independent errors take when each is measured from the mean of thirty passes, -1/(30-1).

A.2 Anchor calibration

The platform returns data in two ways. A monthly table covers all markets in one month, and a full-period series covers one market or state over all 132 months. Each scales its highest value in the request to one hundred. To put all months on one scale, we retrieved full-period series directly for anchors, five markets for the market panel (New York, Los Angeles, Dallas, Atlanta, and Oklahoma City) and five states for the state panel (TX, FL, PA, OH, and CA). We chose them after the first pass on one requirement, search volume above the reporting threshold in all 132 months, so that each anchor gives a complete series. The market anchors range in size from the largest market to a mid-sized one. The anchors agree closely. The logs of their ratio series $T_b(t)/I_b(t)$ (main text Eq. 1) correlate at a mean of .995 for markets and .982 for states.

The following example, with invented numbers, shows how the calibration corrects the monthly tables. The first numbers below are search shares, that is, searches for the seven phrases per million searches of any kind in that market and month. The platform never reports these shares; it reports only 0-100 values computed from them. Suppose that an anchor market had a share of 60 in both January and February, that the top market had 120 in January and 200 in February, and that the anchor's own peak month over the full period had 300. The monthly tables, which scale each month's top market to 100, then report the anchor at 100 x 60/120 = 50 in January and 100 x 60/200 = 30 in February. This is a forty percent fall for a market whose searching never changed, and it arises only because the top market changed. The anchor's own series, on the other hand, reports 100 x 60/300 = 20 in both months, which correctly shows no change. The ratio T/I is 20/50 = 0.40 in January and 20/30 = 0.667 in February, values that equal 120/300 and 200/300, so the anchor's own 60 has dropped out. A second anchor with a peak of 150 gives

0.80 and 1.333, a different level but the same movement, and after we divide each series by its own mean (0.5333 and 1.0667), both read 0.75 in January and 1.25 in February. Real retrievals add sampling noise, so five anchors might read 0.73, 0.75, 0.75, 0.76, and 0.92 in the same month. Their median is 0.75, while the single odd value of 0.92 would pull a mean up to 0.78. For this reason, we use the median. Finally, take a market that is not an anchor and reported 40 in January and 24 in February. After calibration it reads 40 x 0.75 = 30 and 24 x 1.25 = 30, which shows correctly that its searching never changed, although the raw tables suggested a forty percent fall.

A.3 Blind leave-one-out validation of the calibration

To test the calibration on data that it did not use, we dropped each anchor in turn and recomputed the scale factor from the other four anchors. We then calibrated the dropped anchor's values in the monthly tables and compared them with the anchor's own directly retrieved series, which played no part in this calibration. The calibrated values correlate with the held-out series at .997 to .9996 for markets and at .979 to .995 for states, whereas the uncalibrated values reach only .59 to .76 and .14 to .64. Because both series share a long-run shape, a correlation in levels is a weak test on its own, so we also compared the month-to-month changes. In changes, the median correlation is .996 for markets and .984 for states after calibration, but only .42 and .22 before it, so the calibration recovers the monthly movement and not only the shared trend. This test reaches only the anchors, which range from the largest market to a mid-sized one (A.2). For small markets the scale factor is extrapolated from anchors whose sampling behavior differs, so its accuracy there is assumed, not shown. This limit affects the descriptive maps but not the main coefficient, which does not depend on the calibration (A.4).

A.4 Calibration invariance of the panel estimates

As main text Section 3.1 notes, the scale factor becomes a month fixed effect once search enters the models in logs, because it takes one value in each of the 132 months and the same value for every market in that month. In logs, the factor adds the same amount to every market's log search in a given month, and the month fixed effects absorb this amount. When we estimated the main specification on the calibrated and on the uncalibrated panel, both runs gave b = 0.030797. The calibration matters only when we compare levels across months and markets, as in the descriptive figures and maps.

A.5 Reliability and the accumulation of sampling passes

To measure how closely the thirty passes agree, we used Cronbach alpha, which averages all possible split-half coefficients, so that we did not have to choose a particular split (Cronbach 1951). We computed it on the uncalibrated retrievals, with the passes as items and the market-months as subjects, because the sampling noise arises in the retrievals, while the calibration multiplies every pass by the same monthly factor. On the cells observed in every pass (17,162 market-months and 6,532 state-months), alpha is .968 for markets and .997 for states. Cells observed in only some passes lie near the reporting threshold and are not covered by this figure. In this raw form, the mean correlation between passes is .506 for markets and .917 for states, lower than in A.1, where passes are compared within each monthly table. The difference arises because each retrieval scales its own monthly table, and these level differences make up 43 percent of the raw variance. Once the month fixed effects absorb them, in the form that enters the models, log deviations from the month mean, alpha is .997 for markets and .999 for states. The main text reports the lower raw figure.

The main estimate was already stable after the first few passes. We estimated the lag-6 model on the cumulative average of passes 1 through p, with a fixed list of markets, and for p = 1 to 30 the estimate stays between .0277 and .0364 (t = 2.69 to 3.42). Later passes made the measure more reliable without changing the conclusion.

A.6 Censoring and missing data

A blank cell records low search, not a random gap in the data. A cell is blank when the market's search volume that month fell below the reporting threshold in all thirty retrievals, so a blank tells us that searches were at or below the lowest level the platform reports for that market. Blanks are 2,146 of the 27,720 market-months (7.7 percent of cells and 0.86 percent of person-months). They are concentrated in the smallest markets and the earliest years, at 18 percent of cells in 2015 and 4 to 6 percent from 2020 on. In 118 markets no month is blank, and the sparsest market has a value in 14 percent of months. The 19 markets with fewer than 60 usable months hold 0.7 percent of the population and 2.3 percent of the estimation sample. The outcome has no missing values, because a month with no recorded incident is a zero. The main estimate uses the observed cells of all 210 markets, with no screen and no filled values (main text Section 3.4). Imputation models fitted to the reported months, including chained equations, assume that a blank month resembles the reported ones, whereas search in a blank month fell below the

threshold in all thirty retrievals. D.1 therefore shows such fills for contrast only, and they never enter the main estimate. It also reports screens and fills that respect the reporting rule.

A.7 Criterion validity and re-measurement

Background checks recorded by the NICS system are a monthly proxy for gun purchases, but they exist only by state. For this comparison, we counted handgun, long-gun, and multiple checks (the clean definition) and correlated them over time with the state panel of search. The median within-state correlation is .78 across 50 states, where the District of Columbia counts as a state and Hawaii drops out because its clean count does not vary. Across states, search and background checks per capita correlate in each month at a mean Spearman rho of .65. The market panel cannot be compared with the checks directly, so we first aggregated the markets to states. The aggregated series follows the directly retrieved state series closely (median within-state r = .98 across 48 states), and it tracks background checks as closely as the state panel does (median r = .77 across 47 states for both).

The relationship with background checks is not limited to the surge years. Across the 49 states other than the two permit-exempt states (Hawaii and Kentucky), the median within-state correlation is .71 in 2015-2019, .82 in 2020-2021, and .67 in 2022-2025. Part of the co-movement in the pre-surge years and the retreat, however, comes from the shared seasonality of purchasing. Once we remove calendar-month means, the correlation stays at .83 in the surge, while it falls to .48 and .55 in the other two periods. Across the same 49 states, the elasticity of background checks with respect to search is .19 in 2015-2019, .43 in 2020-2021, and .37 in 2022-2025.

We also asked whether search runs ahead of purchases or moves in the same month, because the answer changes how the lead of search over violence should be read. In each state, we correlated log search in month t with log background checks in month t+h, and the median within-state correlation peaks at h = 0 (r = .758). It is a clean single peak that falls off evenly on both sides (.522 at h = -1 and .563 at h = +1). This matches what Brownstein et al. (2020) found in national monthly data, where search and background checks were most highly correlated at zero lag. The peak remains at h = 0, though flatter, after we remove calendar-month means from both series (median r = .685). Search therefore moves with background checks in the same month, not ahead of them. As in main text Section 6.1, we assume that the same holds within markets, where no

series of checks can verify it. On this assumption, the lead of search over violence is an interval from purchase to violence, not from search to purchase.

An independent retrieval of the same phrases in 2022 agrees closely with the present panel by state and moderately by market, and it gives a similar estimate. This retrieval was made in early 2022 with different software and about 200 samples. It correlates with the present panel at a median monthly cross-sectional r of .887 across states and .639 across markets. Within markets, the two retrievals agree more closely over time in larger markets (.751, .636, and .407 from the largest to the smallest size band) and in 2020-2021 than before. The markets and years in which the association appears are therefore not where the two retrievals disagree most. We also put the 2022 retrieval in place of the present panel as the regressor in the baseline specification for 2015-2021, on the same cells (194 markets, N = 14,343). It gives b = .0202 (t = 2.74), against .0240 (t = 2.34) for the present panel. The 2022 series can enter uncalibrated because the month fixed effects absorb the month scale, and no input from the present retrieval touches it.

A.8 Term basket and content check

We set the basket of seven phrases in 2022, when we first collected these data, and have not changed it since, so the phrases were not chosen with the outcome data in hand. The basket has enough volume for a combined index, but not enough to be broken down by phrase, which is why we call the construct acquisition-related search and rest its link to purchasing on the criterion results in A.7.

For the three phrases with enough volume, the searches that accompany them are mostly about buying guns and safes. We checked content validity in the sense of Mellon (2014) by retrieving the related queries of each phrase, the searches that users make alongside it. "Gun sales" returns private, used, and nearby gun sales, and "gun store near me" returns gun shops, pawn shops, and gun stores with an FFL, the federal license that gun dealers must hold. "Gun safe" returns best gun safes, nearby gun safes, safes at named retailers, and safe brands, which are searches for buying a safe. The other four phrases return no related queries at these volumes, and "gun maintenance" has too little volume to pull the basket toward upkeep. For the three phrases whose related queries we can see, those queries are about buying guns or gun safes, which supports the label acquisition-related search.

A.9 Sample-coverage map

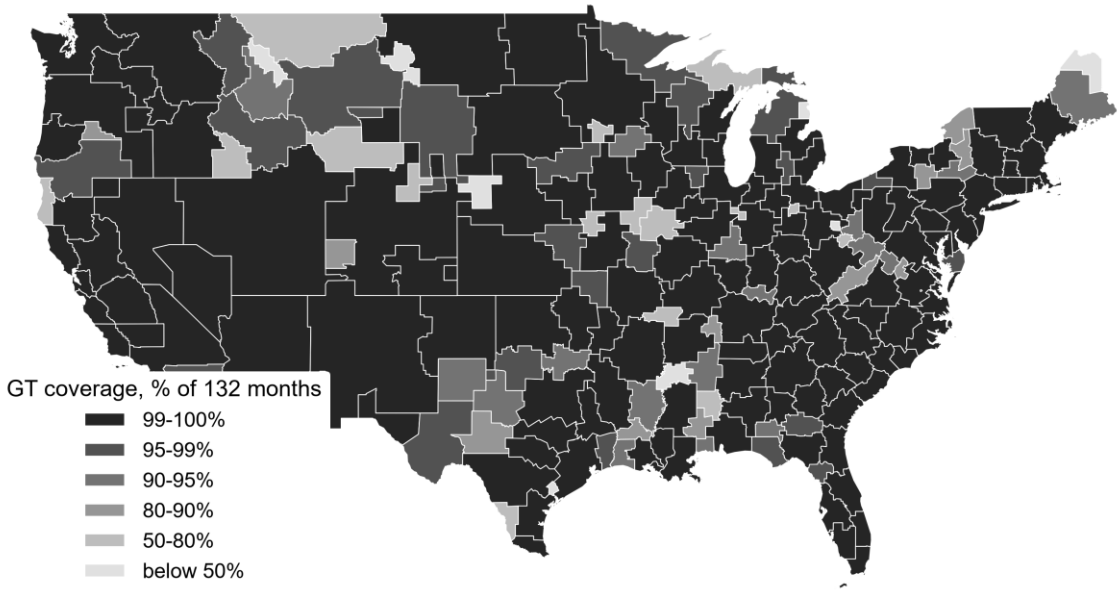


Fig. A1 Coverage of the search panel by media market. The shading gives the share of the 132 study months in which a market has a search value. The map shows the contiguous United States, so the markets of Alaska and Hawaii are not shown. Coverage is lower in smaller markets. The sparsest market has a search value in 14 percent of months, and every market enters the analysis.

Appendix B. Validation of the outcome

B.1 Vital-statistics benchmark

We compared the GVA data with the official death records of NVSS, which code the cause of death from death certificates (ICD-10 X93-X95). By state and month, GVA's count of people killed (victims and suspects, with incidents that were suicides alone left out) correlates with the NVSS count of firearm homicides at a median within-state r of .908 across the 32 states with at least 60 unsuppressed months. The other states cannot be evaluated, because NVSS suppresses counts of 1 to 9. Pooled across the 3,904 evaluable state-months, the counts correlate at r = .984, and at r = .926 after each state's mean is removed. At the market-month level, the GVA fatal rate correlates with the SHR firearm-homicide rate at a pooled r of .717 and a within-market r of .656 across 19,968 market-months without 2021 (B.4). The NVSS regression sample of main-text Table 2 spans 51 states, because every state has at least one unsuppressed month.

We also checked whether GVA's coverage changed in ways that could produce the association. We define the capture ratio as GVA victim deaths divided by NVSS firearm homicides in each state-month, leaving out suspect deaths because NVSS classes them as legal intervention. Its median across the evaluable states is .90. If GVA's coverage had improved over the decade or were better in large states, part of the association could reflect better coverage. The ratio, however, shows no gradient with state size ($t = -1.31$) and no drift that differs by size ($t = 0.62$). Its mild upward trend (+.0028 per year, $t = 1.92$) is far too small to produce the estimates, since the coefficient is roughly eighty times the largest bias that this trend could create. NVSS data for 2025 are provisional and may undercount deaths.

B.2 Same-sample outcome-swap checks

Table B1 The GVA outcome re-estimated on each comparison sample

Outcome source	Level	Sample	N	b (SE)	t
Supplementary Homicide Reports, firearm homicide	market-month	202 markets, 2015-2024 except 2021	18,008	.0324 (.0101)	3.21
Gun Violence Archive, on the homicide-report sample	market-month	202 markets, 2015-2024 except 2021	18,008	.0283 (.0112)	2.53
Vital statistics, firearm homicide	state-month	state-months with an unsuppressed NVSS count (51 states)	4,338	.0747 (.0279)	2.68
Gun Violence Archive, on the vital-statistics sample	state-month	state-months with an unsuppressed NVSS count (51 states)	4,338	.0720 (.0308)	2.34

Table notes: Every row is the lag-6 specification of main-text Eq. (2), with population weights, unit and month fixed effects, and inference clustered on the unit. Each GVA row uses exactly the sample of the row above it. Florida is excluded from the homicide-report rows because the program covers it inadequately, and 2021 is left out of them for the reason given in B.4.

Table B1 re-estimates the GVA outcome on the sample of each comparison source, so that the sample is held fixed and only the source changes. On the 202 markets and the months of 2015-2024 except 2021, the homicide reports give .0324 ($t = 3.21$) and GVA gives .0283 ($t = 2.53$), and the restriction itself moves the GVA estimate little (.0283, against .0308 on the full panel). On the 4,338 state-months with an unsuppressed NVSS count, GVA gives .0720 ($t = 2.34$), up

from .0531 on the full state panel (D.7), so part of the gap between the own-sample rows of main-text Table 2 reflects the evaluable sample. NVSS gives .0747 (t = 2.68) on the same state-months, so the two sources give similar estimates once the sample is held fixed, and the small gap that remains reflects what each source counts.

B.3 Scope of the outcome, and the firearm/non-firearm difference test

We use fatal incidents as the main outcome for two reasons. GVA compiles its incidents from media reports and police sources, which record deaths more completely than injuries, and the two comparison sources record fatal violence only (main text Section 3.2). Nonfatal incidents therefore support no main claim, and we use them only in labeled robustness checks. On the baseline specification, the nonfatal estimate is .0193 (SE .0198, t = 0.97), and fatal and nonfatal incidents together give .0299 (SE .0197, t = 1.52), against .0308 for fatal incidents alone. The nonfatal estimate is smaller than the fatal one, though not significantly so. To test the gap, we regressed the difference between the two transformed rates on lagged search in the same model, and the fatal-minus-nonfatal difference is +.0115 (SE .0173, t = 0.67, p = .51). Within markets the two rates move together only weakly (median r = .12; .36 in the 25 largest markets), consistent with month-to-month noise in the nonfatal series.

For the firearm-specificity test of main text Section 5.4, we compare NVSS deaths from assault by firearm discharge (X93-X95) with deaths from all other assault (X85-Y09 without the firearm codes). If a general rise in violence drove the association, non-firearm homicide should respond as well. We joined the two CDC WONDER final mortality databases (main text Section 3.2), which match exactly where they overlap in 2018-2020 (n = 1,274 state-months, r = 1.0000, mean ratio 1.000). We estimated both outcomes on the same sample, the state-months from July 2015, where the six-month lag starts, through December 2024 (N = 3,705, G = 51). Search predicts firearm homicide with a coefficient of .0784 (t = 2.73), while the coefficient for non-firearm homicide is .0001 (t = 0.01). Stacking the two equations gives a difference of .0783 (SE .0285, t = 2.75), with a state-clustered wild bootstrap p of .024. At 80 percent power, the non-firearm equation could have detected a coefficient of .0207, 26 percent of the firearm coefficient, so a shared confounder that moved the non-firearm rate by that much would likely have shown up in it. On the shorter window of 2015-2020, the coefficients are .0752 (t = 3.60) and .0042 (t = 0.77), and the difference is .0710 (t = 3.68, p = .001), with a minimum detectable effect of .0152 (20 percent). The longer window ends in December 2024 because the non-firearm series ends there.

B.4 Market-level firearm specificity (Supplementary Homicide Reports)

With death records, we can run the firearm-specificity test of Supplement B.3 only by state. NVSS suppresses monthly counts of 1 to 9, and below the state level only the largest counties record enough firearm homicides each month to avoid suppression. The Supplementary Homicide Reports, by contrast, record the weapon in each homicide at a finer geography, so we could repeat the test by market and month. We took the SHR homicides for 2015-2024 (Kaplan 2025, cited in the main text), classified each one by the primary offender's weapon as firearm or non-firearm, and aggregated the incidents from counties to markets as monthly rates on the denominators of the main panel. Florida (eight markets) is excluded, because the SHR covers it inadequately. Apart from the outcome, the model is the baseline specification. We estimate it once for firearm homicide and once for non-firearm homicide and ask whether search predicts the first more strongly than the second. To test the difference between the two coefficients, we stacked the two equations and bootstrapped by market with 999 draws. Each bootstrap p is two-sided, twice the share of draws on the far side of zero.

In the SHR data, we code a market-month without any record as zero, although some of these months may come from agencies that did not file (main text Section 3.2). This coding breaks down in 2021, the first year after the FBI retired its Summary Reporting System (Federal Bureau of Investigation 2021, cited in the main text). Several of the largest markets show a sharp dip in 2021 that disappears in 2022. For example, SHR firearm homicides were 710, 30, and 771 in Los Angeles in 2020, 2021, and 2022, 520, 160, and 440 in New York, and 825, 414, and 636 in Chicago. If we counted the missing reports of 2021 as zeros, they would appear as real declines, so we leave that year out. When we leave out both 2021 and 2022, the firearm estimate is .0320 ($t = 3.11$), and the difference is .0263 ($p = .016$), the same as in the main sample.

The homicide reports repeat the pattern found in the death records. On the main sample of 202 markets ($N = 18{,}008$), search predicts firearm homicide with $b = .0324$ ($t = 3.21$), whereas the coefficient for non-firearm homicide is .0062 ($t = 1.30$). The difference between the two is .0263 (bootstrap $t = 2.61$, $p = .010$). Adding Florida back (210 markets, $N = 18{,}806$) changes little, giving .0332 ($t = 3.34$) for firearm and .0068 ($t = 1.48$) for non-firearm homicide and a difference of .0264 ($t = 2.56$, $p = .006$). The contrast also does not depend on leaving out 2021. With 2021 kept, the firearm estimate is .0450 ($t = 3.43$), the non-firearm estimate is .0090 ($t = 1.60$), and the difference is .0361 ($p < .001$). Within markets, the SHR firearm-homicide rate also moves with

the GVA fatal-incident rate in the same month, with a median correlation of .754 that is positive in all 200 evaluable markets. In 17.6 percent of the 19,968 market-months no SHR homicide is recorded, and the asinh transformation accommodates these zeros.

Appendix C. Outcome construction protocol

The outcome is built from the incident file that GVA supplied, which holds 433,988 incidents for 2015 to 2025, 99.99 percent of them with coordinates. We placed each incident in a 2020 Census county by its coordinates and then in a media market through the county-to-market crosswalk. The crosswalk misses 66 counties. We renamed two of them to their current FIPS codes and assigned the other 64 to the market of the neighboring county with the longest shared border (51 counties) or, failing that, of the nearest county (13 counties). We then applied the exclusions by exact matches of GVA's incident tags. We exclude suicides and keep murder-suicides, which carry interpersonal tags alongside their suicide tags. Exact matching matters here. GVA's tag for fatal incidents reads "Shot - Dead (murder, accidental, or suicide)", and a rule that searched for the word suicide would have removed every fatal incident. Our protocol follows earlier research with GVA data, and its full text, with citations, is in the replication package. Variants that remove officer-involved shootings and defensive gun uses, or add nonfatal incidents, are in D.1 and B.3. The table counts the exclusions in order on the records placed in a market, so that the last two rows equal the sums of the outcome panel.

Step	Rule	Incidents remaining
All records in the GVA file	1 January 2015 to 31 December 2025	433,988
Placed in a media market	coordinates present (28 records lack them)	433,960
Someone was shot	victims or suspects killed or injured, one or more	433,952
Not a suicide alone	drop records tagged as a suicide or suicide attempt and not as a murder-suicide	426,559
Not an unintentional shooting	drop records with any unintentional tag (accidental shooting or discharge, playing with or cleaning a gun, hunting accident, thought gun was unloaded, self-inflicted injury that is not a suicide or suicide attempt, child self-injury, child-on-child shooting, and a child firing a gun that hit a non-child victim)	407,323
Fatal (the outcome)	at least one person killed; the other 249,530 are nonfatal	157,793

Appendix D. Robustness compendium

D.1 Sample-rule, missing-data, and outcome-construction variants

Table D1 lists the samples behind the estimates in the main text and in this supplement, with the rule that defines each. The samples differ because each model needs different data, and where two estimates are compared, we compare them on the same sample.

Table D1 Samples used in the main text and the supplement

Estimate	Where reported	Units	N	Sample rule
Baseline lag-6 estimate	Main text Eq. (2), Tables 2-3	210 markets	23,426	search observed at t and t-6
Lag profile, lag 6	Main text Fig. 2b	210 markets	24,384	search observed at t-6
Twelve-month response	Main text Eq. (3), Table 1	192 markets	19,950	search observed at t through t-12
Twelve-month response with six leads	Main text Table 1; D.11	189 markets	18,556	search also observed at t+1 through t+6
Same, with market-by-calendar-month fixed effects	Main text Table 1; D.11	184 markets	18,520	as above, singleton cells dropped
Seasonal and excess parts of search at lag 6	D.11	206 markets	23,301	search observed at t-6, in a calendar month observed in at least three of the years 2015-2019
Excess part of search, twelve-month response	Main text Section 5.1; D.11	167 markets	18,927	search observed at t through t-12, with every calendar month observed in at least three of the years 2015-2019
Homicide reports (SHR)	Main text Table 2; B.4	202 markets	18,008	2015-2024 except 2021, Florida excluded
Vital statistics (NVSS)	Main text Table 2	51 states	4,338	state-months with an unsuppressed NVSS count
NVSS firearm and non-firearm homicide	Main text Section 5.4; B.3	51 states	3,705	the same state-months for both outcomes, July 2015 to December 2024
The 2022 retrieval in place of the present panel	Main text Section 6.3; A.7	194 markets	14,343	the same cells in both retrievals, 2015-2021

Table notes: N counts market-months for markets and state-months for states. In Fig. 2b, each lag k requires search only at t-k, so N varies by lag; the N shown is for lag 6.

We start from the full panel of 27,720 market-months, the 210 markets observed for 132 months. The first six months of each market have no six-month lag, which removes 1,260 market-

months. Of the remaining 26,460, another 3,034 drop because search is blank at t or at t-6 (958 at t only, 1,142 at t-6 only, and 934 at both), which leaves the 23,426 market-months of the baseline estimate. The sample rule of the lag-6 estimate in main text Fig. 2b asks for search only at t-6, so it keeps 24,384 cells, and on these cells the estimate is .0301 (t = 2.85).

Screening out sparse markets changes the estimate little. Screens at 50, 80, and 90 percent coverage give .0308 (t = 2.86), .0303 (t = 2.77; 184 markets), and .0317 (t = 2.83; 172 markets), close to the .0308 for all 210 markets. Only the markets with complete coverage (118 markets) give a larger estimate, .0407 (t = 3.02). The coefficient barely moves until only the markets with complete coverage remain. If noise near the reporting threshold had created the association, dropping sparse markets would have lowered the coefficient. The markets with complete coverage are also the largest, so the .0407 reflects the stronger association in large markets (main text Section 5.6). The main estimate applies none of these screens and uses all 210 markets.

Filling the blank cells, by contrast, lowers the estimate in every version that we tried. In the first version, which respects the reporting rule, we drew a value of log search for each blank uniformly between a floor and the market's lowest reported value. We completed 50 panels in this way and combined them by Rubin's rules, with cluster-robust variance within each panel. With the floor at half the market minimum, b = .0186 (SE .0077, t = 2.41), and with the floor at half the global minimum, b = .0085 (SE .0041, t = 2.10), on N = 26,460. Because the draws ignore the outcome, they act like extra noise in the regressor and pull the estimates toward zero. What they show is how sensitive the estimate is to a stated floor.

Fills that give every blank one fixed low value lower the estimate for a similar reason. Setting every blank to the market's lowest reported value gives .0210 (t = 2.45), setting it to half of that value gives .0165 (t = 2.36), and setting it to half the global minimum gives .0050 (t = 1.94). Values that do not depend on the outcome, assigned to the regressor in the worst-measured markets, add variance to log search within markets and pull the slope toward zero. For contrast only, we also tried two fills that break the reporting rule, a market median and a linear interpolation of the interior gaps. They give .0291 (t = 2.77) and .0259 (t = 2.49, N = 26,308), and all 1,994 interpolated values lie above the market's lowest reported value, where a blank month should not lie under the reporting rule.

Whether search was reported in the preceding month does not explain the estimate. If blanks depended on violence, whether search was observed would itself predict the outcome. An indicator that search was observed at t-1 has a coefficient of .0180 (t = 1.32) and leaves the search coefficient at .0312 (t = 2.89). With the indicator at t-7, the month before the regressor, the two coefficients are .0082 (t = 0.60) and .0313 (t = 2.84).

We also varied the definition of the outcome, and each of these four versions stays positive and significant. On the per-lag sample, leaving out murder-suicides gives .0306 (t = 2.96), and leaving out single-participant incidents, in which one person was shot and no unharmed suspect was recorded, gives .0209 (t = 2.53). On the baseline sample, removing officer-involved shootings from the fatal outcome (13,476 incidents) gives .0300 (t = 2.90), and removing defensive gun uses as well (a further 6,488) gives .0293 (t = 2.84).

The estimate is larger with population weights than without them. We report wild cluster bootstrap p-values from 9,999 draws here, against 99,999 for the baseline p = .0051. With and without population weights, all 210 markets give .0308 (p = .0061) and .0110 (p = .21). The 118 complete-coverage markets give .0406 (p = .0043) and .0206 (p = .066), and the routine used for the screens above gives .0407 for the same weighted sample. On the per-lag sample of main text Fig. 2, the unweighted estimate is .0098 (t = 1.09). The markets with complete coverage are larger, so the complete-case estimates describe larger markets, not a cleaner sample. Main text Section 5.6 discusses why the weighted and unweighted estimates differ. The caution about small units that led Black and Nagin (1998) and Stephens-Davidowitz (2014) to exclude them is handled here by the weights and by these checks, and no market is dropped.

D.2 Confounder controls

Several other changes of the period could have moved search and violence together, and we added controls for those that we could measure. All estimates in this section use the sample rule of the lag-6 estimate in main-text Fig. 2b (baseline .0301). With state pandemic stringency (OxCGRT) and market mobility (workplaces, residential, and retail) in the model, the estimate moves from .0301 to between .0299 and .0303 (t = 2.85 to 2.92). On the sample that these data cover, it moves from .0540 to .0532 (t = 3.21). On the window that the protest data cover, total protest counts from the Crowd Counting Consortium leave the estimate at .0432, against .0433 without them (main text Table 3), and racial-justice-tagged counts leave it at .0433. With the

racial-justice counts at lags 0, 3, and 6, the estimate is .0404 (t = 3.24) on a smaller sample (N = 17,705), and the protest terms are null. Local unemployment gives .0306 (t = 2.87), and weather gives .0237 (t = 2.32), the largest change among these controls. One possible explanation is that violence follows the seasons, and the weather controls absorb part of each market's own seasonal cycle (main text Section 4). A linear trend for each market gives .0268 (t = 3.00), against .0301 without trends. For law enforcement activity (main text Section 5.5), officer-involved shootings, current and lagged six months, give .0276 (t = 2.73). State-month arrests for weapons, drug, and all offenses give .0303 (t = 2.90), and officer-involved shootings together with weapons arrests give .0281 (t = 2.77). Because the outcome includes officer-involved shootings, the first of these controls overlaps the outcome, but removing those incidents from the outcome changes the estimate little (D.1).

Leaving out the most turbulent months or any one region also keeps the estimate positive. Excluding summer 2020 (May to September) gives .0267 (t = 2.47), and excluding May 2020 through June 2021 gives .0236 (t = 2.20). With unemployment and weather added to the summer-2020 exclusion, the estimate is .0216 (t = 2.05). Leaving out one census division at a time keeps every estimate positive, and the weakest has t = 1.81.

We built the controls from public sources, except the officer-involved shooting counts, which come from the GVA file, and assigned them to markets as follows. We assigned stringency to each market from its primary state, the state that holds the largest share of its population, and we aggregated mobility from counties with population weights. We set both controls to zero in the months that their sources do not cover, which for mobility includes the months after October 2022. The protest control enters only in the months that the source covers, 2017 through 2024, and it is not part of the main specification. We aggregated unemployment (BLS Local Area Unemployment Statistics) to market rates from the county counts. We averaged each county's mean monthly temperature and total monthly precipitation (NOAA nClimDiv) within each market, weighting each county by its population. We also used state-month arrest counts from the FBI Crime Data Explorer. For firearm laws, we counted the restrictive and the permissive laws in force in each month in the RAND State Firearm Law Database (version 6.0, compiled through September 2025) and assigned these stocks to markets by county population share. Main-text Table 3 reports the estimates with the law controls.

D.3 Inference variants

The p-values in the main text come from a wild cluster bootstrap by market. It imposes the null hypothesis and uses Rademacher weights, which flip the sign of each cluster's residuals with probability one half (B = 99,999). This bootstrap gives p = .0051 for the baseline estimate and p-values between .0008 and .033 for the single lags 4 to 9, except the weaker lag 8. For conventional cluster-robust Wald tests, we compare the test statistic with, we compare the test statistic with an F distribution with q and G−1 degrees of freedom, where q is the number of coefficients tested and G is the number of markets. Clustering by market could overstate the precision if markets in the same region share shocks, which is a real possibility here because the identifying variation is regional. We therefore also handled the dependence in the ways listed in Table D2, and the lag-6 estimate stays significant under each alternative standard error and permutation test in the table (market centroids are population-weighted county centroids).

Table D2 Alternative standard errors and permutation tests

Method	Statistic	Value
Market-clustered wild bootstrap (210 clusters)	p	.0051
State-clustered wild bootstrap (48 primary states; Rademacher / Webb)	p	.0191 / .0199
Cumulative 0-12: market / state WCB	p	.0259 / .0277-.0301
Driscoll-Kraay (Bartlett; lag truncation 4; 12)	z	3.40; 3.18
Conley spatial (uniform kernel + within-market Bartlett L=12), 300/500/1000 km	z	4.17 / 4.11 / 3.97
Pesaran CD on two-way-FE residuals	CD (p)	4.32 (p < .0001); mean pairwise rho = +.0037 over 20,557 evaluable pairs
Permutation of search series (market-level reassignment)	p (|b|) / p (|t|)	.0005 / .0065
Permutation of search series (state-block reassignment)	p (|b|) / p (|t|)	.0015 / .0135
Effective number of clusters (Carter et al. 2017), market / state	G^*	96.5 / 22.1
Leave-one-market-out coefficient band	b range	[.0252, .0334]; largest influence LA

The rows address different forms of dependence. State clustering, with the few-cluster weights of Webb (2023), allows for shocks that the markets of a state share. In the permutation rows, we reassign each market's whole search series to another market at random, keeping the outcomes and weights in place, and p is the share of permuted estimates at least as large in absolute value as the observed one. Each permutation test uses 2,000 random reassignments. When we reassign

the series in blocks of whole states, each state receives the series of a randomly drawn state, matched by population rank, and the series of a state with fewer markets are reused in order. The co-movement within states is then kept under the null, and the test does not rely on a large number of clusters. Driscoll and Kraay (1998) designed their errors for dependence across all units in the same period, while the errors of Conley (1999) allow for dependence among nearby units. Both truncate the serial correlation at a cutoff lag, and both give z statistics larger than the market-clustered t, which is one reason we keep the more conservative market-clustered bootstrap for the main inference. Finally, the CD test of Pesaran (2015) pools the 20,557 pairs of markets that share at least 24 months of data, and with so many pairs, it detects even the small mean correlation that remains in the residuals (+.0037).

D.4 Functional form, units, count models, and equivalence

Chen and Roth (2024) showed that coefficients on an asinh outcome change with the units in which the outcome is measured, and our estimates behave in this way. With the fatal rate per 10,000, per 100,000, and per 1,000,000 residents, the per-lag coefficient is .0036, .0301, and .0520, so the main text does not convert any coefficient into a number of incidents. The rates in our panel are also small, with an incident-weighted mean of .579 per 100,000 per month. Bellemare and Wichman (2020) noted that the asinh behaves like a log only for values above about ten. At values as small as ours it is nearly linear, so its coefficient is closer to a change in the level of the rate than to an elasticity.

A linear model of the rate per 100,000, with no transformation, gives the same result. Weighted by population, this model gives .0365 (t = 2.82) at lag 6 and a cumulative 0-12 response of +.117 (t = 2.41).

We also estimated Poisson pseudo-maximum likelihood (Santos Silva and Tenreyro 2006) on incident counts, with a log-population offset, two-way fixed effects, market clustering, and the same sample rule as the baseline specification. These estimates have the same sign as the baseline, and their precision depends on how the markets are weighted. The offset already weights each market's rate residuals by population, so the unweighted Poisson model is the counterpart of the weighted linear model. It gives t = 1.80 at lag 6 and t = 1.50 for the cumulative response. If we also add population weights, each market is weighted by its population squared,

and the largest markets dominate. That version gives a lag-6 coefficient of .1530 (95 percent CI .0175 to .2885) and a cumulative response of +0.302 (t = 1.81).

The association does not appear in whether a month has any fatal incident. A linear probability model of any fatal incident in the month gives t = -1.53. The count among months with at least one incident gives t = 2.27, but this estimate describes a selected sample.

Before 2020, the estimate is statistically equivalent to zero within the size of the pooled estimate. In two one-sided tests with cluster-robust errors, the 2015-2019 coefficient (b = .0023, SE .0094) is equivalent to zero within plus or minus the pooled per-lag estimate (+/-.0301, p = .0018). Equivalence within plus or minus half of the pooled estimate (+/-.0151) is not shown (p = .089). A smaller effect before 2020 therefore cannot be ruled out.

D.5 Heterogeneity detail

Without population weights, the estimate for all markets is small and not significant (b = .0098, t = 1.09). In the largest markets, however, the unweighted coefficient is .066 (t = 3.10) for the top 25, .047 (t = 3.04) for the top 50, and .034 (t = 2.69) for the top 75. Without weights, the coefficient is .066 (t = 3.10) in the 25 largest markets, .047 (t = 3.04) in the 50 largest, and .034 (t = 2.69) in the 75 largest. It is null for the markets ranked 51 to 100 (b = −.0036, t = −0.21) and for those below the top 100 (b = −.0042, t = −0.33). The association is therefore concentrated in the largest markets, and the 50 largest hold 68 percent of the population. In the weighted model, dropping the ten largest markets still leaves b = .0180 (t = 2.13). Among the 50 largest markets, unweighted, the coefficient without 2020-2021 is .042 (t = 2.56).

The null in the mid-sized markets is unlikely to come from sampling noise. For the markets ranked 51 to 100, we compared the calibrated panel with a separate single-pass retrieval of each market's own full-period series (the second retrieval mode of main text Section 3.1, applied to every market). The two measurements correlate at r = .89 after market and month means are removed, and both show no association. After the odd-even correction for sampling error, which uses the average of the even-numbered retrievals as an instrument for the average of the odd-numbered ones (D.6), the estimate for these markets also stays near zero (b = .0056, t = 0.33; N = 6,268). Sampling noise is therefore unlikely to explain the null, although both measurements could share a weakness of the search measure in these markets.

The correction for random sampling error does not close the gap between the periods. In the weighted model on all markets, with the odd-even correction, the 2015-2019 coefficient is .0062 (t = 0.56), and the 2020-2025 coefficient is .0539 (t = 3.93).

Across markets, a larger search surge went with a larger rise in fatal violence. The 2019-to-2020 change in search predicts the 2019-to-2021 change in fatal violence with a slope of +.147 (t = 2.90, population-weighted). Because the changes run over one to two years, the slope measures the response accumulated over that time, in the same way as the twelve-month cumulative response (main text Section 5.1). Fig. D1 shows the timing of the search wave by census division.

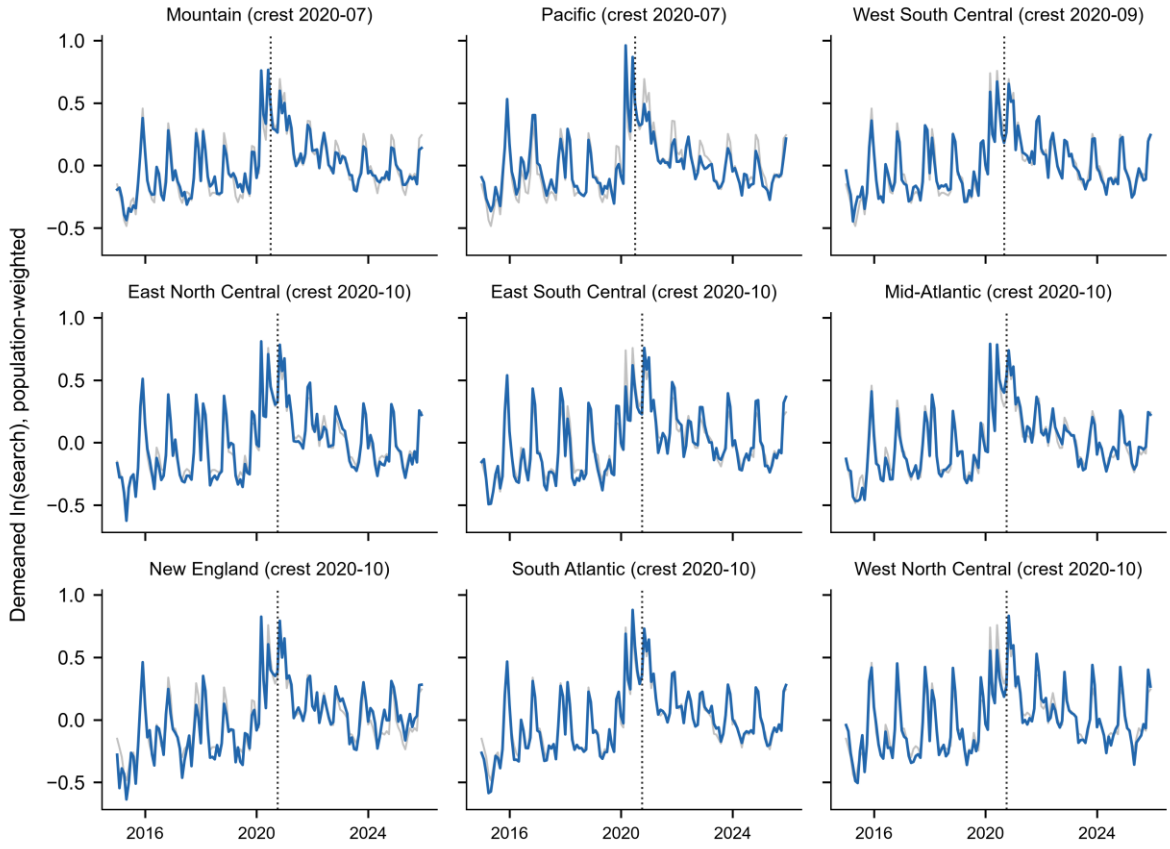


Fig. D1 Wave timing by census division. The blue line is the division's population-weighted, demeaned ln(search), and the gray line is the national series on the same scale. The dotted line marks the division's crest, the peak month of a nine-month centered moving average within 2020-2021, and the panel title shows the crest month. Panels are ordered by the order of crest month.

D.6 Identification anatomy and measurement-error detail

Main text Table 3 (Panel A) narrows the set of markets that each market is compared with, and the estimates show where the identifying variation lies. When each market is compared only with markets of similar size in the same month, the estimate stays significant (t = 2.57). It loses significance when the comparison is limited to markets in the same census division (t = 1.83) or in the same state (b = .0126, t = 1.50). We also restricted the within-state comparison to the 34 states with at least three markets (190 markets), and there the estimate is .0127, with a confidence interval from -.0041 to +.0294, which is inconclusive.

Sampling error in search could weaken the estimate within states, so we corrected for it with instrumental variables. The regressor is the mean of the odd-numbered passes, and the instrument is the mean of the even-numbered passes, in the spirit of the obviously related instrumental variables (ORIV) of Gillen et al. (2019) but without their stacked estimator. Inside the state-by-month comparison, the reliability of the regressor, measured as the slope of the odd-pass mean on the even-pass mean, is .897. The corrected coefficient is .0167 (SE .0102, t = 1.63), and its interval includes both zero and the baseline. With market and month effects only, the corrected coefficient is .0387 (t = 3.27, N = 22,869). The errors of adjacent passes correlate positively (A.1), and adjacent passes fall in different halves, so the instrument shares a little of the regressor's error and the corrected coefficient may still be slightly too small. The correction addresses only classical sampling error, not the censoring of small markets (A.6).

Most of the identifying variation comes from the timing and shape of the wave. We split each market's two-way demeaned search series into an amplitude part, the market's loading on the national wave times that wave, and a timing-and-shape part, the remainder. Over the full panel, the timing-and-shape part carries 89.2 percent of the weighted identifying variance and the amplitude part 11.0 percent (cross-term -0.2 percent), and this split is exact. Inside the surge window, the amplitude share rises to roughly 32 percent, but there the split is only approximate.

D.7 State-level results

To compare our results with earlier studies that used national trends or state data, we also estimated the model on the state panel. The lag-6 coefficient is .0531 (t = 1.90), and the wild bootstrap gives p = .103. The response has the same sign and shape as in the market panel, but it is only marginal at the coarser state level. The two search series align closely, with the market

panel aggregated to states matching the directly retrieved state series at a median within-state r of .98 (A.7).

D.8 Specification curve

Most specifications in the curve give a positive estimate. We estimated 2,052 combinations of outcome definition, lag, weights, sample rule, fixed effects, controls, and transformation (the figure is in the replication package). The lags run from 4 to 9, the band that we chose after inspecting the estimates (D.11). Of the coefficients, 88.9 percent are positive, and 31.2 percent are significant at the 5 percent level (median +.0257). The preferred specification (lag 6, weighted, all markets, two-way fixed effects, no controls, asinh) gives +.0301 and sits at the 57th percentile. Weather controls and lag 8 pull the estimates down the most, and population weights matter as well, since 46 percent of the weighted specifications are significant, compared with 17 percent of the unweighted ones. Outcomes that add nonfatal incidents, as Schleimer et al. (2021) did, give weaker estimates, with medians between +.0112 and +.0123 and only 11.7 to 13.9 percent of these specifications significant. For the fatal outcome alone, the median is +.0306, and 39.2 percent of its 1,440 specifications are significant.

D.9 Frozen-origin forecasting exercise

We also asked whether search helps to forecast fatal violence six months ahead, using only the information available at the time of the forecast. Each month from December 2018 through June 2025 served as a forecast origin, and at each origin we refit every model and recomputed every feature, scaler, and threshold from the data up to that month. We scaled each market's errors by its seasonal-naive error over the training window (Hyndman and Koehler 2006) and compared the scaled squared errors in the style of Diebold and Mariano (1995), with market clustering. Adding search to a panel autoregression lowered the scaled squared error by .0074 (t = 1.50) unweighted and by .0417 (t = 3.93) weighted. The gain comes essentially from one episode, the regime shift of the surge, which was common to all markets, and adding search made forecasts worse for target months from August 2020 through March 2021, before larger gains in mid-2021. Market-clustered t statistics count markets, not independent forecasting episodes, so they overstate how much replicable information search adds. For the gradient-boosting benchmark the comparison is sign-mixed. In the surge window, unweighted, this benchmark does worse with search (t = -2.85), because search moved outside the range seen in training and tree models cannot extrapolate. Over all targets, it gives t = -1.84 unweighted and +2.76 weighted. At our

pre-determined threshold, none of the models functions as a reliable alarm system. Specifically, an alarm is triggered when a forecast surpasses a market's 24-month trailing mean by 20 percent—the same criterion used to identify a true surge (spike) in the realized data. The regression and tree models caught .104 to .165 of the spikes, with false-alarm rates of .078 to .122. The seasonal-naive benchmark, which repeats the value of the same month a year earlier, caught .366, with a false-alarm rate of .284, against a base rate of .322.

D.10 Reverse-path sensitivity

The leads in main text Section 5.5 test direction without conditioning on violence's own past. Granger (1969) proposed asking instead whether one series helps to predict another beyond what the other's own past already predicts, and we apply this idea in both directions. The forward equation (D.10a) regresses violence on search at t-6 and on violence's own lags, and the reverse equation (D.10b) regresses search on violence at t-6 and on search's own lags:

$$\operatorname{asinh}(y_{mt}) = \sum_j r_j \operatorname{asinh}(y_{m,t-j}) + d \ln(S_{m,t-6}) + a_m + g_t + e_{mt} \qquad \text{(D.10a)}$$

$$\ln(S_{mt}) = \sum_j s_j \ln(S_{m,t-j}) + f \operatorname{asinh}(y_{m,t-6}) + a_m + g_t + u_{mt} \quad \text{(D.10b)}$$

Here y_{mt} is the fatal-incident rate and S_{mt} the search index of market m in month t, a_m and g_t are market and month fixed effects, and j runs over the own lags in the table. Both series are standardized to weighted standard-deviation units on one sample (N = 20,672, 200 markets), and we test $d - f$ by stacking the two equations (Rademacher cluster-score wild bootstrap by market, B = 99,999). With a panel of T = 132 months, the bias that fixed effects create in models with lagged outcomes (Nickell 1981) is negligible.

Table D3 Comparing Forward and Reverse Search–Violence Associations

Own-lag set	Forward d (t)	Reverse f (t)	Difference $d - f$ (t)	Bootstrap p
{1}	+.0555 (3.02)	+.0114 (2.34)	+.0441 (2.77)	.003
{1,2,3}	+.0501 (3.05)	+.0102 (2.31)	+.0399 (2.72)	.004
{1,2,3,6}	+.0497 (3.06)	+.0101 (2.30)	+.0395 (2.72)	.004
{1,2,3,6,12} (primary)	+.0477 (3.02)	+.0070 (1.78)	+.0407 (2.64)	.006
{1-6}	+.0500 (3.18)	+.0101 (2.30)	+.0399 (2.83)	.003
{1-6,12}	+.0481 (3.14)	+.0069 (1.78)	+.0411 (2.74)	.004

Across all six sets of own lags, search predicts later violence more strongly than violence predicts later search. The forward coefficient hardly moves (t near 3.0). With short runs of own lags, the reverse coefficient is positive (t near 2.3), but it weakens to t = 1.78 once the twelve-month own lag absorbs the slow seasonal wave in search. In every row the difference $d - f$ is

significant (t = 2.64 to 2.83, p = .003 to .006), although we read it as robust in sign, not as a precise point difference.

We do not view this comparison as definitively settling the direction. One reason is that the two series are measured with very different precision. Search averages thirty retrievals, while the violence rate of a single market in a single month is noisy, and noise in violence weakens both the reverse coefficient f and the forward equation's control for the past of violence. Measurement error alone could therefore make search appear to move first. Our correction for sampling error (D.6) concerns search, not violence, so it does not address this problem, although the forward association survives it (b = .0387 on its own scale, which cannot be compared with the standardized coefficients). The other reason is that the comparison assumes that no firearm-specific shock reached regions at different times and raised acquisition search first and firearm violence later through a separate regional channel.

If violence drove search, search in a market could rise after a local shooting that killed four or more people, but we find no such rise. After the 292 such incidents (287 market-months in 105 markets), we compared search in the affected market with its mean over the six months before the event. The difference is t = -0.02 for months 0 to 2 and t = 0.09 for months 3 to 6. Porfiri et al. (2019) found a media channel from mass shootings to firearm acquisition at the national and state levels, whereas we find no response within the affected market.

D.11 Seasonal decomposition and seasonality-immune designs

To see when each series rises within the year, we described the seasonal profiles of search and of fatal violence in 2015-2019. We used the market-years with all twelve months observed, which cover 170 markets for search and 210 for violence. Within each market-year, we subtracted the yearly mean from ln(search) and from the asinh rate, and then we averaged these deviations by calendar month with population weights. Search was highest in November (+.37) and December (+.35) and lowest in May (-.21). Fatal violence, on the other hand, was highest in July (+.033) and lowest in February (-.054). Since both profiles repeat every year, the search profile correlates with the violence profile when it is moved six to ten months forward (r = .32 to .57), and it correlates by the same amounts when it is moved two to six months back.

Gun purchasing and gun violence both rise and fall over the year, and this shared yearly cycle cannot show which of them moves first. A cycle that repeats every twelve months has the same

value six months before a given month as six months after it, so a lag-6 coefficient on the seasonal part of search could not tell whether search leads or follows violence. We therefore split lagged log search into each market's repeating calendar profile, made of twelve calendar-month means fit to 2015-2019 with at least three observations per cell, and the excess over that profile. Both parts enter the baseline specification together (N = 23,301 in the 206 markets with enough pre-2020 data). We draw no claim about direction from the seasonal coefficient and read only the excess, which can separate lags from leads. The profile is fit to 2015-2019, however, so the excess in those years is partly flattened by construction.

Table D4 Excess vs. Seasonal Search Components Across Periods

Window	Excess b (t)	Seasonal b (t)
2015-2019	-.0092 (-0.74)	+.0423 (+2.36)
2020-2021 (surge)	+.0519 (+2.92)	+.1360 (+4.41)
2022-2025 (retreat)	+.0199 (+1.62)	+.0788 (+3.80)
Full panel	+.0209 (+1.82)	+.0730 (+4.66)

The excess component is significant only in the surge window, and in 2022-2025 it is smaller and not significant. Because the profile is fixed at its 2015-2019 amplitude, however, a larger seasonal swing during the surge would appear in the excess component and not in the seasonal one. The year-over-year design below is the check against this reading.

In the year-over-year design, every stable seasonal profile drops out. A seasonal swing that grew during the surge would survive the differencing. However, a twelve-month cycle looks the same six months ahead and six months behind, so such a swing would tend to appear at the six-month lead as well as at the lag. We regressed the twelve-month change in the asinh rate on the twelve-month change in log search six months earlier, with market and month fixed effects and population weights. Over the full panel, b = +.0081 (t = 0.90). In 2020-2021, the forward coefficient is +.0337, positive but not significant at the 5 percent level (t = 1.87), while the six-month lead, our placebo, stays near zero (-.0048, t = -0.24). The other periods show no forward association, with -.0043 (t = -0.35) in the routine years of 2016-2019 and -.0072 (t = -0.52) in the retreat of 2022-2025. Since differencing doubles the error variance of the regressor, some attenuation is expected. The lead stays at zero during the surge while the forward coefficient is positive, and this is the pattern that matters for the seasonal question.

We also removed seasonality from both series directly, in two ways. Market-by-calendar-month fixed effects give .0210 (t = 1.76), against a same-sample baseline of .0301, and subtracting the

pre-2020 profiles from both series gives .0170 (t = 1.50), against .0306. The estimates fall because these designs remove the periodic share of the variation, and they are also less precise, with confidence intervals that include both zero and the baseline. These designs therefore cannot tell the two apart.

In the raw profile, the association concentrates at lags 4 to 9. In the full sample of main-text Eq. (3), lags 4-9 of the raw regressor are jointly significant (p = .0001), and lags 0-3 and 10-12 are not (p = .31). We chose this band after inspecting the estimates.

In the excess component, the cumulative response remains, but it is spread over the year. We entered all thirteen excess lags together, as in Eq. (3), on the 167 markets with a seasonal pattern, from at least three of the years 2015-2019, for every calendar month (N = 18,927; none of the smallest markets qualifies). The cumulative 0-12 response is +.1088 (t = 2.59), against +.1005 (t = 2.40) for the raw regressor on the same sample. A repeating profile summed over thirteen lags is nearly constant within a market, and the market fixed effects remove it, so the cumulative response is almost immune to the seasonal component. The timing is a different matter. In the excess component, lags 4-9 are jointly significant (p = .0149), but the other lags are marginal, not null (p = .0819), and lags 0, 9, and 12 carry much of the weight (t = 2.35, 3.54, and 2.34). Part of the sharp 4-9 band in the raw profile therefore comes from the seasonal phase. Split by lag, about half of the excess response of +.1088 falls on lags 0-3 (+.0528), and the rest falls on lags 4-9 (+.0466) and lags 10-12 (+.0095). The three parts add to .1089 because of rounding (the per-lag values are in the table below).

Table D5 Per-Lag Coefficients for Excess and Raw Search Regressors

Lag	Excess b (t)	Raw b (t)
0	+.0219 (2.35)	+.0135 (1.69)
1	+.0137 (1.51)	+.0011 (0.13)
2	+.0081 (0.99)	-.0064 (-0.85)
3	+.0091 (1.17)	-.0012 (-0.15)
4	+.0169 (1.97)	+.0195 (2.29)
5	+.0043 (0.59)	+.0206 (2.87)
6	+.0044 (0.50)	+.0185 (2.11)
7	+.0057 (0.78)	+.0172 (2.36)
8	-.0112 (-1.25)	-.0046 (-0.53)
9	+.0265 (3.54)	+.0307 (4.07)
10	-.0044 (-0.54)	-.0079 (-0.99)
11	-.0046 (-0.69)	-.0073 (-0.95)
12	+.0185 (2.34)	+.0068 (0.90)

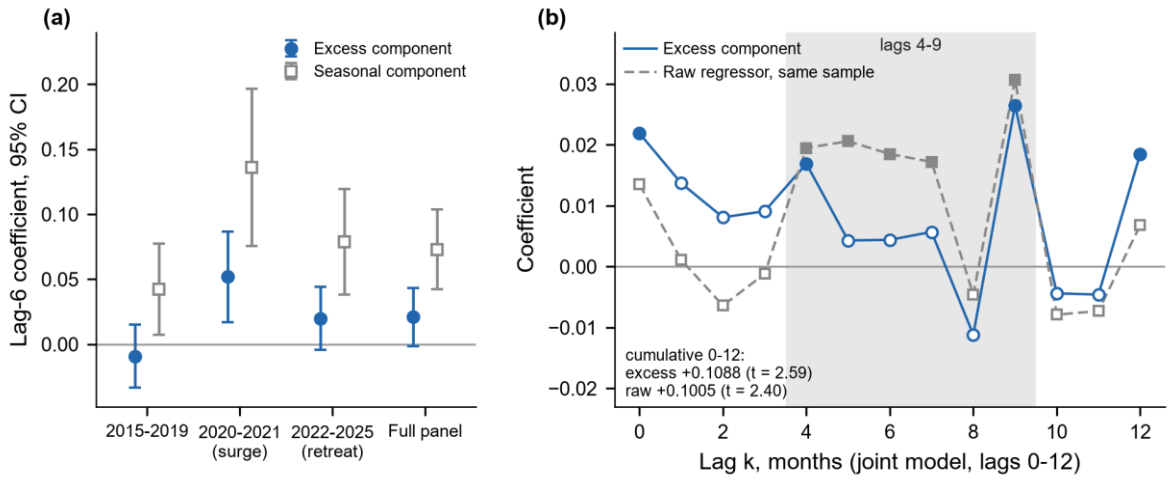


Fig. D2 The frequency split of lagged search. Panel (a) shows the lag-6 coefficients for the excess (filled circles) and seasonal (open squares) components, by period and for the full panel, with 95% market-clustered CIs. The values are in the first table of this section. Panel (b) shows the per-lag coefficients from the joint distributed-lag model on the common sample (N = 18,927, G = 167) for the excess component (circles, solid line) and the raw regressor (squares, dashed line). In panel (b), filled markers are individually significant at the 5 percent level, and the shaded region marks lags 4-9. The joint test of lags 4-9 gives p = .0149 for the excess component and p = .0001 for the raw regressor, and the joint test of the other lags (0-3 and 10-12) gives p = .0819 and p = .2352, respectively. The cumulative 0-12 response is +.1088 (t = 2.59) for the excess component and +.1005 (t = 2.40) for the raw regressor.

Part of the lead pattern appears to come from the same yearly cycle. Taken one at a time, the leads t+1 to t+8 are all positive, with t = 2.00 to 3.38 for t+3 to t+7, and the leads from t+9 to t+12 are close to zero (t+1 to t+6 are plotted in main text Fig. 2b). The leads near six months are about as large as the lags near six months (.0319 at t+5 and .0278 at t+6, against .0301 at lag 6). When the six leads of main text Table 1 are added to the joint model of Eq. (3), they are not jointly significant (p = .38), but their sum is +.043 (SE .022, t = 1.97), and the lags 0-12 sum to +.074 (t = 1.89; N = 18,556 in 189 markets). We then added market-by-calendar-month fixed effects to the month effects. These remove the seasonal pattern of both series within the regression, so that each series is compared with its own market's usual level in the same calendar month. On nearly the same sample (N = 18,520 in 184 markets), the six leads then sum to +.022 (t = 0.94; joint p = .58), no single lead is significant, and the lags 0-12 sum to +.088 (t = 2.32).

With the excess component of search in place of the fixed effects, the six leads sum to +.019 (t = 0.87) and the lags to +.093 (t = 2.38; N = 17,778 in 167 markets, with search observed from t-12 to t+6).

Twelve leads span a full year, so a fixed yearly cycle tends to cancel from their sum. We entered leads t+1 to t+12 and lags 0-12 in a single model on the market-months in which search is observed from t-12 to t+12. For raw search, the twelve leads sum to +.018 (t = 0.61; joint p = .10) and the lags to +.082 (t = 2.08; N = 17,316 in 185 markets). Adding market-by-calendar-month fixed effects gives a lead sum of +.010 (t = 0.34; joint p = .94) and a lag sum of +.095 (t = 2.39; N = 17,293 in 183 markets). For the excess component, the leads sum to +.012 (t = 0.41; joint p = .76) and the lags to +.097 (t = 2.47; N = 16,697 in 167 markets).

On the same samples, the twelve leads of raw search entered without the lags are jointly significant (p = .011), although their sum (+.057, t = 1.49) is not. The twelve leads of the excess component entered without the lags sum to +.063 (t = 1.64; joint p = .50). The thirteen lags entered without the leads sum to +.094 (t = 2.14) with raw search and +.104 (t = 2.36) with the excess component. Because search persists from month to month, and raw search also repeats each year, leads entered without the lags partly stand in for them, so the test of direction uses the models with both.

Appendix E. Reproducibility, audit, and design decisions

E.1 Pipeline and audit

The replication package includes an audit manifest. For each number reported in the paper, the manifest names the script that produced it, the command that reruns it, the output file, and the input data. The bootstrap seeds are fixed, so a rerun reproduces the outputs to the digit.

E.2 Analyses run and excluded, with the exclusion rule

Some analyses that we ran are not reported in the paper, and we excluded them under a single rule that does not depend on the sign of their results. The rule, applied after estimation, excludes an analysis when no main claim rests on it and it is not stable enough to bear weight, because of a small effective sample, marginal power, or a design defect found in testing. Under this rule we excluded three event-study designs for firearm laws, because their power was marginal and announcement dates were not available. A difference-in-differences that grouped markets by the

early surge was excluded because offsetting differences in the composition of the groups left the net effect uninterpretable. An interrupted time series was excluded because it cannot separate events that hit the whole country at once. Under the same rule we also set aside three other designs. A decomposition into rising and falling search had fragile interaction terms, a split by domestic violence was unstable, and the first stage of the designs based on ownership exposure failed. The same rule removed a design that related winter search to the following summer's violence, because the same winter also predicted the summer before it. Finally, we excluded a figure that aligned the timing of the wave market by market, because the timing varies too little across markets to test alignment one market at a time. The pooled models nonetheless draw most of their identifying variation from differences in that timing and in the shape of the wave (D.6). None of the excluded analyses produced evidence in the opposite direction of the main association, and all remain in the archived record with their outputs.

Results that the main text summarizes in a sentence or does not report are in the sections above. The figures of the specification curve and of the forecasting exercise and the remaining detail of each check are in the replication package.

E.3 Numbers vintage

Every search value in the main text and this supplement comes from the thirty-pass retrieval completed on 23 July 2026, except the 2022 comparison retrieval in A.7 and the single-pass full-period retrievals in D.5, and every rate uses the same Census county population estimates. Years through 2019 use the vintage-2019 county estimates, and years from 2020 on use the vintage-2025 estimates. We first defined the analysis sample with an 80 percent coverage screen (184 markets). After estimation, we found that the screen did not drive the estimate (.0303 with it and .0308 without it; D.1), and we widened the sample to all 210 markets.

E.4 Data availability

The replication package includes the query specifications that reproduce the CDC WONDER exports. At the time of writing, the Harvard Dataverse record for the county-to-market crosswalk (Sood 2016) showed only its citation page, and the file could not be downloaded. The package therefore includes the copy we downloaded earlier and the county-to-market table used in every script.

References cited only in this supplement

Bellemare MF, Wichman CJ (2020) Elasticities and the inverse hyperbolic sine transformation. Oxf Bull Econ Stat 82:50-61. https://doi.org/10.1111/obes.12325

Black DA, Nagin DS (1998) Do right-to-carry laws deter violent crime? J Legal Stud 27:209-219. https://doi.org/10.1086/468019

Carter AV, Schnepel KT, Steigerwald DG (2017) Asymptotic behavior of a t-test robust to cluster heterogeneity. Rev Econ Stat 99(4):698-709. https://doi.org/10.1162/rest_a_00639

Conley TG (1999) GMM estimation with cross sectional dependence. J Econom 92:1-45. https://doi.org/10.1016/s0304-4076(98)00084-0

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253-263. https://doi.org/10.1080/07350015.1995.10524599

Driscoll JC, Kraay AC (1998) Consistent covariance matrix estimation with spatially dependent panel data. Rev Econ Stat 80:549-560. https://doi.org/10.1162/003465398557825

Gillen B, Snowberg E, Yariv L (2019) Experimenting with measurement error: techniques with applications to the Caltech cohort study. J Polit Econ 127:1826-1863. https://doi.org/10.1086/701681

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424-438. https://doi.org/10.2307/1912791

Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. Int J Forecast 22:679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Nickell S (1981) Biases in dynamic models with fixed effects. Econometrica 49:1417-1426. https://doi.org/10.2307/1911408

Pesaran MH (2015) Testing weak cross-sectional dependence in large panels. Econom Rev 34:1089-1117. https://doi.org/10.1080/07474938.2014.956623

Porfiri M, Sattanapalle RR, Nakayama S, Macinko J, Sipahi R (2019) Media coverage and firearm acquisition in the aftermath of a mass shooting. Nat Hum Behav 3:913-921. https://doi.org/10.1038/s41562-019-0636-0

Santos Silva JMC, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88:641-658. https://doi.org/10.1162/rest.88.4.641

Simonsohn U, Simmons JP, Nelson LD (2020) Specification curve analysis. Nat Hum Behav 4:1208-1214. https://doi.org/10.1038/s41562-020-0912-z

Webb MD (2023) Reworking wild bootstrap-based inference for clustered errors. Can J Econ 56:839-858. https://doi.org/10.1111/caje.12661