Check for
updates

Proxies for legal firearm prevalence

Jessica Jumee Kim[1] · Kenneth C. Wilbur[2]

Abstract

Product acquisition policies define legal markets. Policy evaluations require data but prevalence data are not always available. We introduce Legal Firearm Prevalence (LFP), a direct behavioral measure based on the population of firearm licensees in Massachusetts, and argue that it can help evaluate firearm sales and usage restrictions. LFP is not directly measurable in most firearm markets, so we test candidate proxies for LFP in several common research designs, finding that firearm acquisitions are the best proxy in every research design tested. We update the classic study of guns and crime by Cook and Ludwig (2006), finding that choosing an invalid proxy can lead to false research conclusions. We recommend systematic collection and reporting of firearm acquisition data to improve firearm research and inform firearm policy.

1 Introduction

Governments regulate acquisition and use of potentially harmful products, including firearms, alcohol, tobacco, explosives and others. Acquisition policies define legal sales by allowing or prohibiting buyers, sellers and products. Governments also

✉ Jessica Jumee Kim
   jessicajumeekim@gmail.com

   Kenneth C. Wilbur
   kennethcwilbur@gmail.com

1 Rady School of Management, University of California San Diego, 9500 Gilman Drive, #0553, La Jolla, CA, USA

2 Professor of Marketing and Analytics, Sheryl & Harvey White Chair in Management, Rady School of Management, University of California San Diego, Wells Fargo Hall, Room 3W117, 9500 Gilman Drive, #0553, La Jolla, CA, USA

Ⓐ Springer

deter illegal sales and use through penalties and enforcement. For example, firearm acquisition policies include buyer age minimums, background checks and waiting periods; seller license requirements; and product bans on new machine guns, assault weapons and low-quality handguns. Usage policies include concealed carry, open carry and self-defense laws. Deterrence policies include detection, prosecution and punishment of illegal firearm acquisitions.[1]

Buyer and seller behavior interact with government policies to define legal and illegal product markets. We distinguish between three potentially measurable constructs in the firearm context. Total Firearm Prevalence (TFP) is the proportion of the population with access to a firearm. Legal Firearm Prevalence (LFP) is the proportion of the population with legal access to a firearm. Illegal Firearm Prevalence (IFP) is the proportion of the population with illegal access to a firearm. By definition, $LFP + IFP \geq TFP$.

TFP, LFP and IFP are difficult to measure in the United States. We presume that illegal firearm buyers or sellers would conceal IFP to evade detection and punishment. IFP is interesting but we do not study it. We focus instead on LFP, which by itself is difficult to measure.[2] We are unable to find any direct measure of LFP in prior literature.

Researchers urgently need data to study firearm acquisition and usage policy effects on LFP, among other outcomes. A comprehensive research synthesis by Smart et al. (2020) concludes that "current efforts to craft legislation related to guns are hampered by a paucity of reliable information about the effects of such policies." The stakes are high. Federal Bureau of Investigation data show that firearm murders increased faster in 2020 than any previous year on record, up 29% from 10,537 in 2019 to 13,620 in 2020. Gun Violence Archive data show that incidents with four or more gunshot victims increased 127% over six years, from 269 in 2014 to 611 in 2020. US Centers for Disease Control data show that firearm suicides increased 28% over ten years, from 18,735 in 2009 to 23,941 in 2019. Opinion surveys show that beliefs about firearm policy effects diverge along political lines: self-identified Democrats and Democratic-leaning independents favor stricter gun laws, whereas Republicans and Republican-leaning independents favor looser gun laws; both groups tend to predict their preferred policy would reduce crime (Research 2021).

Prior research has seldom distinguished between TFP, LFP and IFP. Instead, lacking data, researchers have typically used proxies to stand in for TFP when estimating policy effects. The most frequent proxy for TFP is Firearm Suicides divided by Suicides (FSS), also known as Percent of Suicides committed with Guns (PSG).[3] FSS measurements depend on mortality data from state vital

---

[1] Castillo-Carniglia et al. (2018) document qualitative differences between firearm acquisition and illegal firearm deterrence policies. Law enforcement officials in two states announced that they would not enforce new Comprehensive Background Check regulations that had been passed by state legislatures.

[2] In fact, the Firearm Owners Protection Act of 1986 restricts the U.S. Treasury Department from "the establishment of any system of registration of firearms, firearm owners, or firearm transactions."

[3] FSS was introduced by Cook (1979) and subsequently validated by two highly cited studies (Azrael et al. 2004; Kleck 2004) as the highest cross-sectional correlate of survey measures of firearm prevalence.

registration offices, so FSS is comparable across geography and time. Kleck (2015) found that 12 of the 19 studies of guns and crime published since 2000 used FSS as a proxy for firearm prevalence.

In this paper we introduce Legal Firearm Prevalence (LFP), a direct behavioral measure based on the population of firearm licensees in Massachusetts. We focus on LFP because firearm acquisition policies directly influence LFP by defining legal markets for guns. Additionally, firearm use policies may motivate or demotivate legal firearm acquisitions. Valid proxies for LFP are needed to better interpret extant research and to enable policy evaluations of how firearm acquisition and usage policies affect LFP in jurisdictions where sales and prevalence data are not available.

1. How well do FSS and other suicide-based proxies explain LFP, based on Massachusetts data from 2010-2017?
2. How well do retail-based proxies explain LFP? We evaluate three candidate proxies: the population of legally registered firearm acquisitions; the population of firearm sales on a leading online platform; and the population of federally licensed firearm retailers. Guns are durable goods with an estimated 393 million firearms in circulation in the U.S. (Small Arms Survey 2018), so it is an empirical question how strongly firearm acquisitions or other contemporaneous retail proxies would correlate with LFP.
3. How do state/month FBI background check data compare to LFP and candidate proxies?
4. How similar or different are LFP and proxy estimates in applied research? We repurpose the classic design of Cook and Ludwig (2006) and compare LFP's effect on homicide to proxy estimates.

The main findings are as follows. Firearm acquisitions are the best proxy for LFP in all research designs tested. Online sales also perform well in most designs. FSS serves as a good proxy for LFP in cross-sectional designs but not in intertemporal or panel designs. FBI data correlate highly with LFP across time at 0.41, but are not available at the county level and therefore cannot be tested in cross-county research designs. Finally, the Cook & Ludwig (2006) exercise shows that choosing the wrong proxy may cause a Type I error by underestimating parameter uncertainty.

We undertake these analyses with three specific goals. First, we hope that the results will help policymakers to evaluate the degree to which published firearm policy evidence which proxies for TFP may also apply to LFP. Our findings indicate that cross-sectional research that uses FSS likely proxies well for LFP, whereas intertemporal or panel research designs do not. Second, we hope that the results can help to inform future researchers in their selection of proxies for LFP. Third, we hope that the results can help motivate policymakers to collect and publish privacy-compliant measurements of LFP or firearm acquisitions, as

that would enable more and better firearm policy-relevant evidence. We discuss several specific actions that are currently available to various policymakers.

Before proceeding, we specify five important caveats to minimize any potential misunderstandings. First, we focus on Massachusetts data because it was the only state that made the requisite LFP information available. Second, we do not claim to estimate causal effects; all findings are descriptive or correlational in nature. The focus of the current article is measurement, not identification.[4] Third, we do not claim to measure or study IFP or TFP. Those topics interest us, and certainly may be affected by firearm acquisition and usage policies, but they are not the subject of this paper. Fourth, we claim no ability to use LFP to evaluate illegal firearm deterrence policies. We suspect IFP or TFP proxies would be needed to evaluate deterrence policy effects. Fifth, we do not advocate for particular firearm acquisition or usage policies; our goal is to facilitate empirical evaluations of firearm policy effects on LFP.

The next section provides the paper's intended contributions. After that, we introduce and describe LFP. The following two sections describe the candidate retail-based and suicide-based proxy variables and compare them to LFP in six common research designs. Subsequently, we apply the research design of Cook and Ludwig (2006) to recent data to compare LFP and proxy estimates, holding data and methods constant. The paper concludes with implications for researchers and policy makers and the limitations of the exercise.

1.1 Relevance and intended contributions

We take an expansive view of the links between policy, quantitative marketing and economics. Any empirical economic analysis of regulated markets must condition on the acquisition policies that define legal transactions in those markets. Marketers operating in regulated markets can influence policy through a variety of actions, including directly via lobbying, or indirectly via demarketing, public relations or product design.

We think the findings make several contributions to quantitative marketing and economics. Evaluating durable goods' market potential requires estimation of market saturation, which in turn requires estimation of durable goods' served available market. To our knowledge, this is the first research to show that durable good acquisitions reliably indicate overall levels of a durable good's penetration. Prevalence and sales are directly related by construction, yet prevalence may also be affected by migration, divestiture or death. Additionally, collectors' purchases may increase sales without increasing overall prevalence. Surprisingly, we find that the number of product retailers does not reliably predict durable good prevalence. We further believe this is the first study of archival firearm purchase data in the marketing literature. Finally, some of the methods we employ might help to guide proxy evaluation in other marketing contexts.

More broadly, we hope to contribute to the policy evaluation literature that spans several fields, including economics, marketing, medicine, public health and public

---

[4] Causal effects of firearm acquisitions are interesting and important, e.g. Donohue et al. (2019).

policy. We offer the first evaluation of proxies for LFP, as opposed to TFP, and introduce two new retail-based proxies for LFP. We also find that FSS is highly correlated with LFP cross-sectionally, but changes in FSS do not reliably indicate changes in LFP. This latter finding may be relevant to several dozen published studies which have used FSS as a proxy for TFP in cross-temporal research designs. Finally, we provide the first assessments of FBI background checks as a proxy for LFP.

2 Legal firearm prevalence

We first discuss firearm prevalence measures and proxies in previous studies, then introduce and describe Legal Firearm Prevalence (LFP). An important distinction is that existing firearm proxies were intended to indicate Total Firearm Prevalence (TFP), which combines legal firearm prevalence with illegal firearm prevalence.

2.1 Extant firearm prevalence measures and proxies

Azrael et al. (2004) and Kleck (2004) compared candidate proxies for TFP to surveys of firearm prevalence. The two surveys used were the General Social Survey (GSS) and the Behavioral Risk Factor Surveillance System Survey (BRFSS). GSS is a biennial national interview-based survey with a response rate of 77% (Kleck 2004). However, the high cost of the interview format limited the national sample size to about 3,000 respondents per year (Azrael et al. 2004), too few to reliably estimate firearm prevalence at the state level. BRFSS is a telephone survey conducted by state health departments with a median response rate of 67% and a median of 2,061 respondents per state per year (Azrael et al. 2004; Powell et al. 1998).

One might ask whether survey data can measure firearm prevalence directly, without need for a proxy. Most literature has not taken that approach. Surveys often feature small samples, design inconsistencies, and insufficient frequency to estimate granular studies of firearm policy effects. Further, firearm surveys may feature non-random response rates, question ambiguities, and respondent misreporting, considering the political and personal sensitivity of the topic.

Prior research has used a variety of behavioral proxies for TFP. Duggan (2001) uses *Guns & Ammo* magazine circulation data to proxy for firearm prevalence. Azrael et al. (2004) and Kleck (2004) evaluated a range of firearm proxies including concealed carry permits, NRA memberships, crime and arrest data, unintentional firearm deaths, outdoor magazine subscriptions and federal firearm licensees. Lang (2013), Briggs and Tabarrok (2014) and Vitt et al. (2018) use FBI background checks as a proxy for prevalence. Kovandzic et al. (2011; 2013) use outdoor sports magazines subscriptions, percentage of those voting Republican in the 1988 presidential election, and numbers of military veterans as instruments for their proxy FSS.[5]

---

[5] Recent NRA magazine subscriptions have flattened (Gilson 2021) despite a four-fold increase in FBI background checks between 2006 and 2020, so we do not suspect that print magazines still offer a good proxy for firearm prevalence. NRA membership data is not publicly available (Ingraham 2021).

A variety of earlier papers have studied firearm license data. Krug (1967), Stolzenberg and D' alessio (2000), and Haas et al. (2007) interpreted concealed carry permits as a proxy for firearm prevalence. Fisher (1976) uses data on handgun licenses and registrations in Detroit to measure firearm availability. Bordua (1986) uses Firearm Owners Identification Cards in Illinois to measure firearm prevalence. Sloan et al. (1988) use concealed-weapons permits issued in Seattle and restricted-weapons permits in Vancouver in their comparison of crime data between the two regions. McDowall (1991) uses handgun purchase licenses issued in Detroit, in addition to FSS. Moody et al. (2010) and Bangalore and Messerli (2013) use gun registration rates to proxy for gun ownership across nations. Hunting licenses have been used to proxy for firearm prevalence and augment other firearm proxies (Andrés and Hempstead, 2011; Kleck 2004; Siegel et al. 2014a; 2014b). Unfortunately, hunting license data are often limited by locally imposed quotas and "therefore may have limited use in identifying changes in firearm ownership over time." (Schell et al. 2020, p. 17) Hunting license quotas may be more related to local animal populations than changes in firearm prevalence. Although firearm licenses have been studied previously, we do not know of any prior work that distinguished TFP from LFP or evaluated proxies for LFP.

Kleck (2015) documents the extensive use of proxies for firearm prevalence in the literature studying links between guns and crime. He counts 19 proxies used by 41 published studies, but concludes that "none of the proxies used in prior research, including [FSS], have been shown to be valid for purposes of judging trends over time."

Schell et al. (2020), motivated by some of the same observations as us, estimate a latent state/year index of household firearm ownership combining survey and proxy data. We view this as an important effort and a worthy goal. Yet one limitation is that measures of ground truth are needed to establish the validity of latent empirical constructs. This concerns us because the Schell et al. (2020) index estimates Massachusetts firearm penetration to have fallen by about one third between 2010 and 2016 (from 12.2% in 2010 down to 9% in 2016), whereas the state's license data showed that the proportion of individuals with valid firearm licenses increased by about half from 2010 to 2016, from 4% to 6%. A second concern is that the latent index uses FSS intertemporally, which has been questioned by Kovandzic et al. (2013) and Hayo et al. (2019), among others.[6]

2.2 Legal firearm prevalence

We measure Legal Firearm Prevalence (LFP) using firearm license data from Massachusetts. All adults in Massachusetts are legally required to maintain an active license in order to purchase or possess a firearm. Firearm licenses expire after six years. New licenses cost $100, as do license renewals.

---

[6] Cook and Ludwig (2019) replied to Hayo et al. (2019) that, as the ratio between relatively rare events, FSS requires sufficient data on population and time to stabilize and limit measurement error; but they do not prove that FSS is a valid intertemporal proxy for firearm prevalence.

We observe all new firearm licenses and license renewals issued between January 1, 2006 and December 31, 2017. For each record, we observe the license issue date, type, status (new or renewal), licensing authority (typically a local police agency), and an anonymous licensee identifier. We use the term "license" to include all firearm licenses that enable legal firearm purchase and possession in Massachusetts. The data consist mainly of Class A Licenses (90%) and Firearm Identification Cards (9%).

We map each licensing authority into the county that contains it. We then measure LFP by counting firearm licensees with one or more active licenses in each county in each month. Counting licensees, rather than licenses, avoids double-counting people who hold multiple valid licenses concurrently. Web Appendix A describes potential left-censoring of LFP in 2010 and 2011 and how we resolve the issue.

2.3 LFP description

Figure 1 shows that the number of firearm licensees increased by 74% over an eight-year period, rising monotonically from 3.9% of the state population in January 2010 to 6.8% in December 2017. The rate of growth was nearly constant, with minor accelerations in 2013 and 2016.

Figure 2 shows the time series of LFP per capita within each of the fourteen Massachusetts counties. Growth rates vary across counties, but downturns are remarkably rare. Franklin County had the highest average LFP, reaching 14.1% in December 2017. The counties with the highest initial LFP in 2010 grew most during the next 7

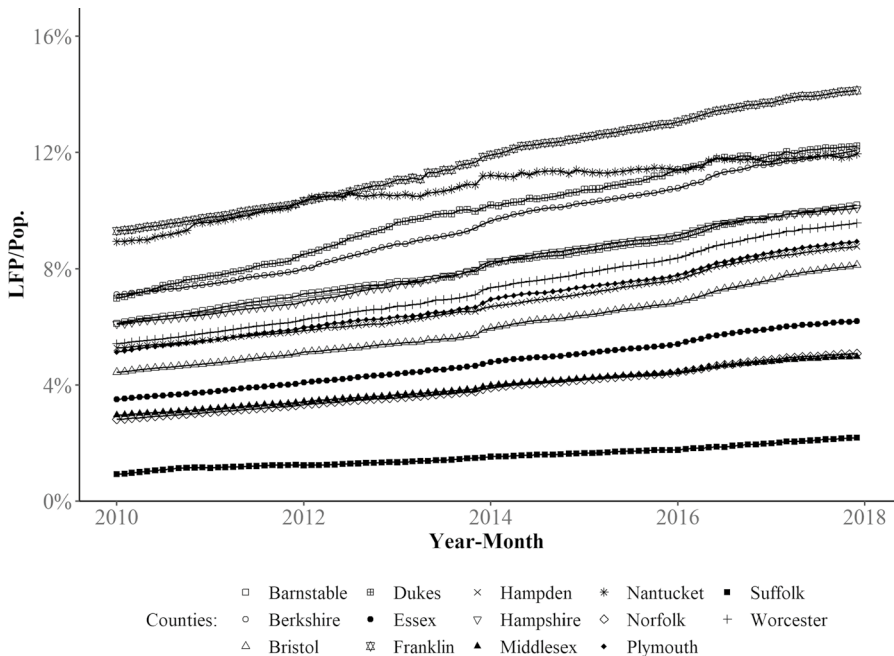**Fig. 1** LFP. Note. LFP indicates the number of active licensees/pop.
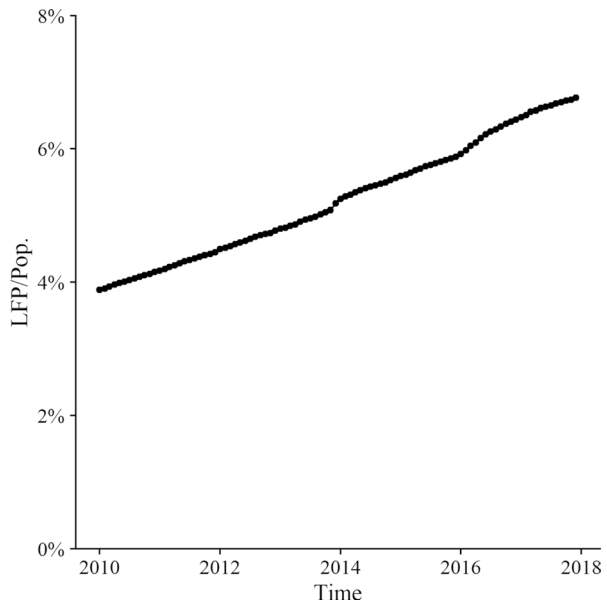
Fig. 2 LFP by County. Note. This figure shows the LFP trend for each MA county.

years, with the exception of Nantucket. Suffolk County had the lowest average LFP, growing from 0.9% to 2.2% over 8 years.

2.4 Strengths and weaknesses of LFP

LFP has several strengths. Most importantly, it counts all firearm licensees based on official government records. It is a population rather than a small or self-selected sample drawn from a population. The fact that licenses must be renewed regularly at substantial cost suggests that LFP should meaningfully reflect both new acquisitions and divestitures of legal firearm access. Unlike survey data, it does not rely on firearm owners' self-reports, and therefore is not subject to sampling error, nonresponse bias, survey design problems or respondent misreporting. Finally, the measure counts distinct people, rather than individual licenses issued or individual firearms, an important property when one individual may hold multiple license types and collect multiple firearms.

The measure also has some weaknesses. Most importantly, we only have data from Massachusetts. Although we filed similar Freedom of Information Act (FOIA) requests with three other states (California, Connecticut and Hawaii), none responded by providing comparable data, despite our repeated inquiries. Second, LFP does not measure firearm lethality, concealability, or concentration of weapons among households or individuals.

We do not claim that LFP also indicates illegal firearm prevalence; hence the term "Legal" in Legal Firearm Prevalence. One of the primary reasons to conduct firearm research is to inform firearm policy, and LFP is a direct outcome of firearm acquisition policies. Therefore, if we seek to understand the effects of policy changes on legal firearm acquisitions, we should distinguish LFP from illegal firearm prevalence. We view measurement of IFP and evaluation of IFP proxies as important topics for future research.

3 Candidate firearm proxies

We analyze two types of candidate proxies for firearm prevalence: retail-based proxies and suicide-based proxies.

3.1 Retail-based proxies

Firearm retailing activity leads directly to firearm acquisitions, suggesting that retail-based measures may proxy for legal firearm prevalence. However, retail-based proxies have two limitations. First, no national database of firearm acquisitions exists, so direct, comprehensive measures of firearm retailing activity are not available for all areas. Second, any retail-based data source will exclude acquisitions that occurred prior to the data sample. Firearms are durable goods, so retail-based data sources cannot provide a fully accurate measure of firearm prevalence. Still, we have found three sources of retail-based data that offer candidate proxies for LFP.

Massachusetts is one of a few states that requires firearm acquisitions to be registered with the state. Acquisition data include Sales, Transfers and Registrations ("STR"). We received the population of Massachusetts STR along with the firearm license data by filing a FOIA request.

We also obtained data on the population of firearm sales intermediated by an anonymous online firearm retail platform (*Online Sales* or "OS"). The data report individual transactions and include buyer zip codes, which we use to assign each firearm acquisition to buyer counties.[7] OS data are less comprehensive than the STR data, as they do not report offline sales transactions through traditional firearm retailers, but they do have two advantages. First, OS data cover the entire United States, suggesting that OS data may provide a retail-based proxy with national scope. Second, some firearm sales might not be legally registered with the state, and therefore excluded from the STR data; or they might not be registered in a timely fashion.

The third retail-based proxy we analyze is the population of Federal Firearms Licensees (FFLs) that sell firearms to consumers, a retail-based proxy that was first considered by Kleck (2004). The data were downloaded from the Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF) website.[8] We matched each FFL address to the county that contained it using ArcGIS, a popular geocoding software.

---

[7] Some zip codes span county borders. In cases where 85% of a zip code's population resides in one county, we assign transactions to that larger county. Otherwise, we drop the transaction.

[8] We only count FFL types 1, 2 and 7 as these indicate retail firearm sales.

The FFL proxy counts the number of firearm retailers in each county in each month and shows substantial intertemporal variation. Licenses for FFLs expire after three years. Substantial numbers of FFLs enter and exit during the sample period.

3.2 Suicide-based proxies

The proportion of firearm suicides, FSS, was validated as the highest cross-sectional correlate of TFP in survey data. We have not found any clear theoretical justification in the literature for FSS as a proxy for firearm prevalence, but FSS is the most frequent proxy for total firearm prevalence in the scientific literature, and is therefore an important proxy to evaluate. We also analyze the proxy's two components, Firearm Suicides (FS) and Suicides (S), as they are readily available and less volatile over time. We lack a compelling theoretical reason to consider S or FS valid proxies; we consider them because they each contribute to variation in FSS.

We count suicides by county and month using Multiple Causes of Death (MCOD) files from the National Center of Health Statistics (NCHS) division of the Centers for Disease Control and Prevention (CDC). MCOD files report information collected from U.S. death certificates provided by state vital registration offices.[9]

3.3 Sample period

All candidate proxies described above are available monthly from January 2010 through December 2017, except for FFL data. FFL data are available from January 2014 through December 2017, with two months of data missing in September and October of 2015.[10]

3.4 Cumulative and static measures

Firearms are durable goods, so most consumers who acquire firearms retain them for substantial periods of time. However, the proxy measures listed above are counts of transitory events (e.g., suicides, firearm acquisitions). Therefore, we consider two forms of each proxy: *static* and *cumulative*. The static version is the count observed within each county and time period. The cumulative version is the accumulation of all activity observed in a county from the beginning of the sample up until a given time period.[11] For FFLs, the contemporaneous value of the proxy is the net accumulation of firearm retailers from the start of the sample until the current time period; therefore, we take the static version of the proxy as the *change* in FFLs from the previous time period to the current time period.

---

[9] We define firearm suicides using standard International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes X72-X74. Suicides are assigned to counties using associated Federal Information Processing Standards (FIPS) codes.

[10] ATF staff were unable to explain why those two months were missing or to provide data prior to 2014.

[11] For STRs, the sample period begins in January 2006. For OS and suicide-based proxies, the sample period begins in January 2010. For FFLs, the sample period begins in January 2014.

We see a meaningful distinction between cumulative retail-based proxies and cumulative suicide-based proxies. LFP is the totality of accumulated firearm acquisitions and divestitures, so accumulated firearm acquisitions should meaningfully reflect some variation in LFP. However, prior literature provides no similar theoretical connection between LFP and suicide data. We include FSS and its components because FSS is frequently used in existing literature, not because we have a theory that predicts suicide-based proxies will reflect LFP. We consider cumulative versions of the suicide-based proxies because we believe they have not been evaluated previously, and in fact we find they perform better than FSS in cross-temporal research designs.

3.5 Summary of proxy data sets

Table 1 summarizes candidate firearm proxies, labels, data sources, sample periods, and descriptive statistics.

4 Methods and results

Previous research has linked firearm proxies to societal outcomes using a wide variety of research designs. Outcome data vary in their geographic and temporal resolutions, such as county/state or month/year. Research designs also vary, such as cross-sectional, time series, panel, etc. We assess the empirical performance of each candidate proxy in four common research designs.[12]

1. Cross-sectional correlation analysis.
2. State/month trends analysis.
3. County/year panel regression.
4. County/month panel regression.

Following Azrael et al. (2004) and Kleck (2004), cross-sectional and intertemporal designs are evaluated with bivariate correlation coefficients. We use regressions with unit and time fixed effects to evaluate candidate proxies in panel designs.

Assessment of candidate firearms proxies in multiple settings may offer two types of insight. One is to inform researchers about the potential utility of a particular proxy within a specific research design of interest. The other is to indicate patterns in order to gain more general insights into the nature of the empirical relationships between each proxy and LFP, in hopes of finding consistent patterns that might apply in other research designs that we do not consider.

---

[12] See, e.g., Moody and Marvell (2005), Cook and Ludwig (2006), Siegel et al. (2013) or Depetris-Chauvin (2015), among many others.

Table 1 Summary of candidate firearm proxies

Proxy	Labels	Source	Sample Period	Summary
Proportion of Firearm Suicides	FSS	Multiple Causes of Death (MCOD) data from the National Center of Health Statistics (NCHS), Centers for Disease Control (CDC)	2010-2017	Mean=21.6%
Firearm Suicides	FS	MCOD data from NCHS, CDC	2010-2017	1,049 total
Suicides	S	MCOD data from NCHS, CDC	2010-2017	4,861 total
Sales, Transfers, and Registrations	STR	Massachusetts Firearms Records Bureau	2010-2017	915,196 total
Online Sales	OS	Leading Online Firearm Retail Platform	2010-2017	24,375 total
Federal Firearm Licensees	FFL	Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATF)	2014-2017, excluding Sept.-Oct. 2015	615 unique FFLs

Table 2 Cross-sectional correlations between LFP and candidate firearm proxies

Proxy	ID	1	2	3	4	5	6	7
LFP/Pop.	**1**	1.00						
FSS	**2**	0.74*	1.00					
FS/Pop.	**3**	0.68*	0.95*	1.00				
S/Pop.	**4**	0.66*	0.83*	0.93*	1.00			
STR/Pop.	**5**	0.86*	0.52	0.41	0.41	1.00		
OS/Pop.	**6**	0.65*	0.10	-0.02	0.01	0.68*	1.00	
FFL/Pop.[‡]	**7**	0.56*	0.39	0.01	-0.11	0.75*	0.43	1.00

*p < 0.05. 14 observations (1 per county) are used to calculate each correlation. [‡]:Data available from 2014-2017, with Sept., Oct. 2015 missing.

4.1 Cross-sectional correlation analysis

We report cross-sectional correlations between candidate proxies and LFP per capita, following Azrael et al. (2004) and Kleck (2004), using population data from the 2010 Census for the 14 counties in Massachusetts. Unlike subsequent research designs, cross-sectional analysis using the full sample does not allow for distinctions between cumulative and static versions of candidate proxies.

Table 2 shows the cross-sectional correlation matrix of LFP and the candidate proxies using the full sample period. All measures except FSS are per capita. All six suicide-based and retail-based proxies are significantly correlated with LFP, ranging from .56 to .86. STR is the most highly correlated with LFP at .86.

FSS is the second most strongly correlated with LFP at .74. This correlation lies within the .64-.92 range reported by Kleck (2004, Tables 1 and 3) and below the .81-.93 range reported by Azrael et al. (2004, Table 3).[13]

The remaining four proxies have correlations with LFP ranging from .56 (FFL) to .68 (FS). The suicide-based proxies are strongly correlated amongst themselves (.83-.95) by construction, as Suicides include Firearm Suicides, and FSS is the ratio of the two. The retail-based proxies correlate with each other to a lesser degree (.43-.75). Correlations between the two sets of proxies are not significant and some are close to zero, suggesting that the retail-based proxies may offer information that is not available in the suicide-based proxies.

4.2 State/Month trends analysis

We rely on intertemporal correlations and visual interpretation to assess the utility of each candidate proxy for LFP in trends analysis. Recall that Fig. 1 shows that

---

[13] Kleck's cross-sectional correlation ranges are from various analyses using multiple GSS constructs and different temporal and geographic units. Azrael, Cook, and Miller's ranges are from different survey measures of gun ownership.

Table 3 County/year regressions

x		Proxy-only Model				Model with Proxy and Controls			
					F-test			Partial	F-test
	Obs.	β_1	t-stat	R^2	p-value	β_1	t-stat	R^2	p-value
FSS	109	-1349.3	-0.02	.000	.986	-12684.5	-1.32	.013	.209
Cum. FS/Cum. S	112	32665.6	0.30	.009	.770	-4916.1	-0.21	.012	.836
FS	112	2018.4	13.50	.480	.000	-178.7	-2.15	.001	.051
Cumulative FS	112	356.6	14.27	.620	.000	189.8	8.45	.912	.000
S	112	449.4	14.93	.559	.000	125.2	2.24	.178	.043
Cumulative S	112	64.1	12.27	.507	.000	31.0	8.20	.842	.000
STR	112	2.5	23.96	.899	.000	1.4	21.21	.839	.000
Cumulative STR	112	0.4	10.59	.738	.000	0.2	12.41	.971	.000
OS	112	83.4	10.79	.785	.000	21.7	4.21	.510	.001
Cumulative OS	112	13.9	8.11	.646	.000	6.1	7.69	.904	.000
Change in FFL[‡]	42	-553.7	-0.74	.002	.471	136.2	0.55	.002	.592
FFL[‡]	56	659.2	14.04	.866	.000	-18.4	-0.08	.001	.937

Each row is a regression of Firearms Licensees (y_{ct}) on Proxy (x_{ct}) weighted by 2010 Census Population with/without County Fixed Effects (α_c) and Year Fixed Effects (λ_t) where c ∈ MA Counties and t ∈ Year := {2010,...,2017}. Standard errors are clustered by county. F-stat p-values test the null hypothesis that the model without the proxy fits the same as the model with the proxy. [‡]:Data available from 2014-2017, with Sept., Oct. 2015 missing.

LFP per capita increased monotonically over time with variable growth rates. In this analysis, LFP and the proxy measures are again measured per capita, except static and cumulative FSS.

Figure 3 shows how static and cumulative suicide-based proxies change over time at the state/month level. The three static proxies are highly variable and their time trends do not track LFP. Suicides correlate with LFP most highly at .28; FSS shows the lowest correlation with LFP at .03.

The cumulative suicide-based proxies are somewhat different. Cumulative FS and S both increase, by construction, and therefore both correlate very highly with LFP at 0.997, although their average growth rates exceed that of LFP. Cumulative FSS, in contrast, correlates negatively with LFP at -.57.

Figure 4 shows how static and cumulative retail-based proxies change over time at the state/month level. The three static proxies are again highly variable. STR correlates with LFP mostly highly at .76, followed by OS at .47; both are statistically significant. Change in FFL correlates negatively with LFP at -.26.

The cumulative retail-based proxies are again somewhat different. Cumulative STR and Cumulative OS both increase monotonically, by construction, and both correlate with LFP at 0.998, although their aggregate growth rates are again larger than LFP. FFL, in contrast, increases slowly but then levels off, and correlates with LFP at 0.73.

Overall, we find that 4 out of 12 candidate proxies correlate very highly with LFP intertemporally, above .997. Among the static proxies, STR and OS are the most
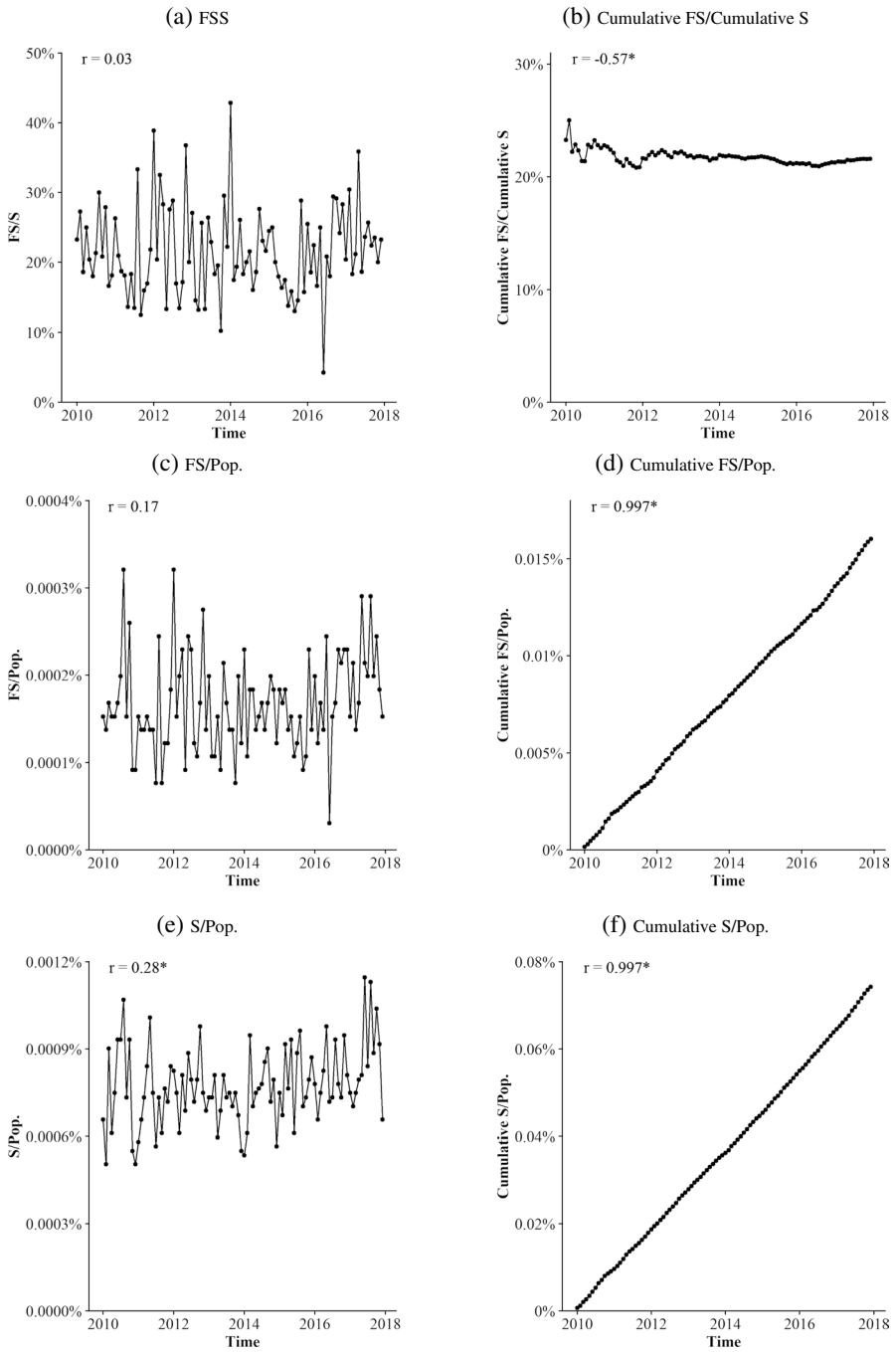
Fig. 3 Suicide Proxies Over Time. Note. r is the Pearson correlation score of the proxy and LFP. *p < 0.05.

promising with correlations of .76 and .47. Static FSS correlates with LFP at just .03, and cumulative FSS is negatively correlated with LFP at -.57.

Web Appendix B reconciles the high cross-sectional correlation and low cross-temporal correlation between FSS and LFP. It repeats the cross-sectional correlation and graphical analyses across four successive two-year subsamples. It shows that FSS is the most volatile of the candidate proxies, thereby explaining its low intertemporal correlation with LFP. Kovandzic et al. (2013) and Cook and Ludwig (2019) treat similar topics in greater depth.

4.3 County/Year panel regressions

Cross-sectional and intertemporal correlations are elegant and valid ways to estimate empirical relationships, but aggregated data may conceal unmeasured confounds that influence both firearm prevalence and firearm proxies. Next, we consider county/year panel regressions, both with and without county fixed effects and year fixed effects to control for unobservable county-specific and time-specific variables.[14]

We do not use per capita measures in the panel regressions, because county populations are only measured directly in the decennial census. The Census provides interpolated population estimates, but those figures are projections rather than direct measurements, and therefore subject to forecasting and interpolation errors. Further, scaling both dependent and independent variables in a regression by common factors can induce spurious correlation (Kronmal 1993; Hayo et al. 2019). Instead, we follow Azrael et al. (2004) in using county populations to weight observations and control for heteroskedasticity. We also cluster standard errors at the county level, following the experimental design rationale described by Abadie et al. (2017).

We specify and estimate two models:

$$LFP_{ct} = \beta_0 + \beta_1 x_{ct} + \epsilon_{ct} \tag{1}$$

$$LFP_{ct} = \beta_0 + \beta_1 x_{ct} + \alpha_c + \lambda_t + \epsilon_{ct} \tag{2}$$

The Proxy-only Model in Eq. 1 is a regression of LFP in county c in year t on an intercept and one candidate proxy denoted x_{ct}. The Model with Proxy and Controls in Eq. 2 augments that simple specification with county- and year-specific fixed effects. We investigate both models because some proxies are more highly correlated with the control variables than others. Models with control variables offer more stringent tests of candidate proxies, as unit and time fixed effects typically explain large fractions of variation in panel data sets.

There is no widely agreed-upon set of criteria to establish proxy validity in panel settings. In the context of firearm prevalence, Azrael et al. (2004) and Kleck (2004)

---

[14] Three county/year observations exhibit zero suicides, leading to undefinable static FSS, so we drop those three cases when estimating the models. FFL proxy models lose observations for periods in which FFL data were not available from ATF.
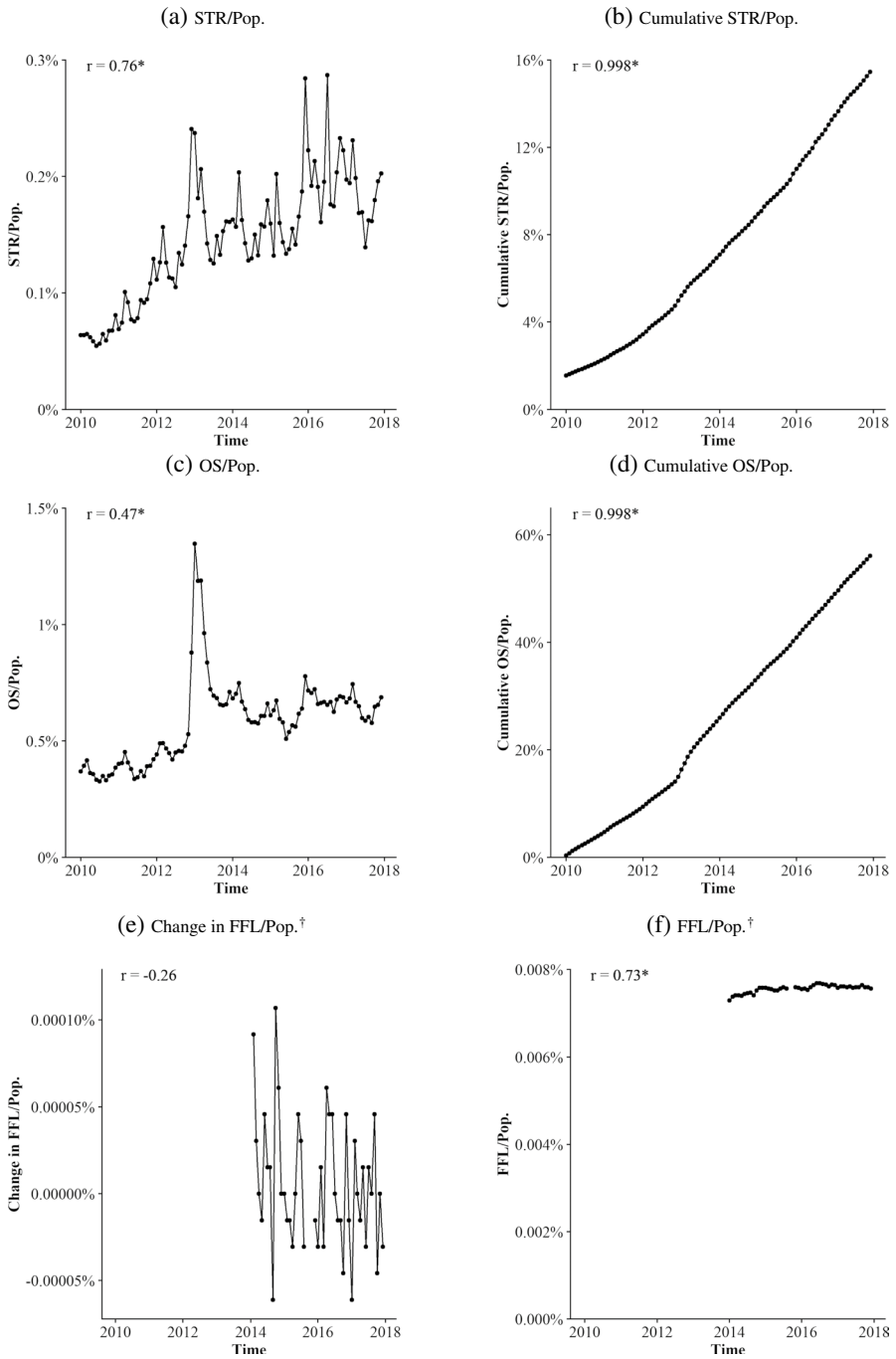
(a) STR/Pop.

(b) Cumulative STR/Pop.

(c) OS/Pop.

(d) Cumulative OS/Pop.

(e) Change in FFL/Pop.†

(f) FFL/Pop.†

Fig. 4 Retail Proxies Over Time. Note. r is the Pearson correlation score of the proxy and LFP. *p < 0.05.‡:Data available from 2014-2017; Sept., Oct. 2015 missing.

focus on size and statistical significance of correlations between candidate proxies and firearm prevalence estimates. Kovandzic et al. (2013, p.484) further emphasize that any valid proxy should explain a large portion of the variance in firearm prevalence. We evaluate the utility of each proxy in each research design using the following criteria:

1. F-test to test proxy inclusion in the model specification.
2. The squared partial correlation coefficient (Partial R^2) to indicate the proportion of variance of LFP that is explained solely by the proxy after partialing out the controls.[15]
3. Estimated proxy parameter sign, magnitude and precision. A perfect proxy would have an effect size equal to one, meaning LFP and the proxy show a 1:1 relationship.

Later in the paper, we also pay attention to the consistency of proxy performance across research designs, as such patterns may suggest proxy utility in untested designs.

Table 3 summarizes the results of both the Proxy-only Model and the Model with Proxy and Controls. Each row of Table 3 reports results from two distinct regressions pertaining to that particular row's candidate proxy. The entire table summarizes 24 distinct regressions.

In the Proxy-only Models, 9 of the 12 candidate proxies are statistically significant predictors of LFP. These 9 proxies explain between 48-90% of the variation in LFP; all have positive relationships with LFP. Three proxies explain less than 1% of the variation in LFP: static FSS, cumulative FSS, and static FFL.

Next we focus on the more stringent Models with Proxy and Controls. 6 candidate proxies each explain more than 50% of the residual variation in LFP. Those 6 include two suicide-based proxies—Cumulative Firearm Suicides and Cumulative Suicides—and four retail-based proxies: Static and Cumulative STR, and Static and Cumulative OS. The t-stat and F-test p-values for these 6 proxies also are significant, indicating the proxy adds important information beyond that provided by the county and year fixed effects.

The remaining 6 proxies explain 0.1-17.8% of the residual variation in LFP after partialing out the controls. Static and Cumulative FSS each explain about 1% of the residual variation in LFP. This echoes the results of Kovandzic et al. (2013), who also find near-zero intertemporal correlations between FSS and survey measures of firearm prevalence.

The Static Suicides proxy is statistically significant and explains 17.8% of the residual variation in LFP. It performs substantially better than FSS, but also

---

[15] It would be easy to misread the Partial R-square statistics reported below, as a high Partial R-square between a cumulative proxy and LFP might be incorrectly interpreted as following directly from both variables' cumulative nature and mutual positive trends over time. In fact, the panel models with controls include time period fixed effects which are completely differenced out before the Partial R-square statistics are calculated. Therefore each Partial R-square statistic solely reflects panel covariation between LFP and proxy remaining *after* unit and time period fixed effects are estimated.

substantially less well than the 6 candidate proxies with Partial R^2 statistics exceeding 0.5.

Three candidate retail-based proxies seem especially promising, with Partial R^2 statistics ranging from .839 to .971, and effect sizes falling within an order of magnitude of 1: STR (1.4), Cumulative STR (0.2) and Cumulative OS (6.1).

4.4 County/Month panel regressions

More granular temporal resolutions allow for better controls for unmeasured confounds while reducing potential aggregation biases. Many societal outcomes (e.g., crime) are measured monthly and therefore can be studied in higher-resolution data that enables more extensive controls. Therefore we test the Models with Proxy and Controls again using county/month panel data. We specify and estimate the following models:

$$LFP_{ct} = \beta_0 + \beta_1 x_{ct} + \alpha_c + \lambda_t + \epsilon_{ct} \tag{3}$$

$$LFP_{ct} = \beta_0 + \beta_1 x_{ct-1} + \alpha_c + \lambda_t + \epsilon_{ct} \tag{4}$$

The models regress LFP in county c in year/month t on an intercept (β_0), one candidate proxy (x_{ct}), a county fixed effect (α_c) and a year/month fixed effect (λ_t). In the second model, we replace the contemporaneous value of the candidate proxy x_{ct} with its first lag, x_{ct-1}. We evaluate first lags of candidate proxies because many longitudinal analyses in long panels (e.g., Cook and Ludwig 2006; Duggan 2001; Khalil 2017) have used lagged proxies as informal checks for reverse causation or simultaneity.

Table 4 shows the county/month panel regression results. Each row in the table reports two unique regressions pertaining to that particular row's proxy: one for the contemporaneous value and one for the first lagged value. All regressions included county and year/month fixed effects.

The qualitative conclusions in Table 4 are remarkably similar to the Proxy and Controls models estimated using county/year data. The qualitative results are nearly identical whether using contemporaneous or lagged values of the candidate firearm proxies.

Among the suicide-based proxies, only Cumulative Firearm Suicides and Cumulative Suicides explain large amounts of residual variance in LFP after partialing out the unit and time controls. The remaining contemporaneous and lagged suicide-based proxies explain 0.0-2.2% of the residual variance in LFP. The only major quantitative difference between county/month results and county/year results is that the residual variance in LFP explained by Static Suicides falls from 17.8% in annual data to about 2.0% in monthly data.

Among the retail-based proxies, Static STR, Cumulative STR, and Cumulative OS perform particularly well, with high Partial R^2 statistics, statistically significant beta coefficients, and effect sizes within one order of magnitude of unity. Cumulative STR again explains the most variation in LFP (98.2%). The two largest quantitative differences between county/month results and county/year results is that the

Table 4 County/month regressions

x	Model with Proxy and Controls					Model with Lagged Proxy and Controls				
	Obs.	β_1	t-stat	Partial R^2	F-test p-value	Obs.	β_1	t-stat	Partial R^2	F-test p-value
FSS	1075	−1131.7	−1.92	.000	.078	1065	−1059.6	−1.90	.000	.080
Cum. FS/Cum. S	1326	−3444.1	−0.34	.003	.739	1312	−3153.8	−0.32	.003	.757
FS	1344	−135.9	−1.97	.000	.070	1330	−117.9	−1.73	.000	.108
Cumulative FS	1344	176.6	9.60	.909	.000	1330	176.9	9.67	.909	.000
S	1344	143.6	2.58	.020	.023	1330	143.3	2.73	.022	.017
Cumulative S	1344	30.0	9.10	.844	.000	1330	30.2	9.05	.843	.000
STR	1344	9.7	20.92	.523	.000	1330	9.7	20.80	.524	.000
Cumulative STR	1344	0.2	12.93	.982	.000	1330	0.2	12.80	.981	.000
OS	1344	79.4	3.56	.191	.003	1330	77.5	3.54	.187	.004
Cumulative OS	1344	6.2	8.84	.919	.000	1330	6.3	8.81	.919	.000
Change in FFL[‡]	616	−174.9	−1.62	.003	.129	602	−179.1	−1.70	.003	.112
FFL[‡]	644	12.8	0.09	.004	.928	630	21.4	0.14	.005	.889

Each row shows two unique regressions of Firearms Licensees (LFP_{ct}) on either Proxy (x_{ct}) or Lagged Proxy (x_{ct-1}) pertaining to that particular row's proxy. Controls include County Fixed Effects (α_c) and Year/Month Fixed Effects (λ_t) where c ∈ MA Counties and t ∈ Year/Month := {2010-Jan,...,2017-Dec}. The regressions are weighted by population and standard errors are clustered by county. F-stat p-values test the null hypothesis that the model without the proxy fits the same as the model with the proxy. [‡]:Data available from 2014-2017, with Sept., Oct. 2015 missing.

Partial R^2 of Static STR falls from 83.9% in annual data to 52.3% in monthly data, and the Partial R^2 of Static OS falls from 51% in annual data to 19.1% in monthly data. These changes in variance explained align with the theory that monthly data allow better controls for unmeasured confounds than annual data.

4.5 FBI background checks

This section seeks to quantify associations between LFP, candidate proxies, and a frequent state/year-month proxy for firearm prevalence, FBI Background Checks. We treat FBI Background Checks separately from other candidate proxies for two reasons. First, Lang (2013) showed recently that FBI background checks correlate highly with GSS survey measures of firearm prevalence in a panel of census divisions and years. Second, FBI Background Checks are only available at the state/ month level, and therefore inevaluable in most of the research designs considered previously.

The Brady Handgun Violence Prevention Act of 1993 required that any person who wants to buy a gun from a federally licensed firearms retailer must submit for a background check conducted by the Federal Bureau of Investigation (FBI). The FBI publishes the number of background checks conducted for firearms purchasers in each state in each month. Most purchasers complete their purchase

shortly after the FBI background check, suggesting that FBI background check data reliably indicate firearm purchase intention. Consequently, recent literature has used FBI background checks to proxy for firearm prevalence (e.g., Briggs and Tabarrok 2014; Lang 2013; Vitt et al. 2018).

Still, FBI background check data have some weaknesses. First, like STR, FBI background checks are a flow variable rather than a stock variable. Second, the FBI only publishes background check data at the state/month level, meaning it is not possible for external researchers to access more granular variation in FBI background check data. Third, the FBI publishes the total count of background checks sought rather than the number of background checks passed or the number of people who seek background checks. That means that a single person could account for multiple FBI background checks in a single state/month, and that FBI data count checks that do not lead to firearm purchases alongside those that do lead to firearm purchases (FBI 2017b). Fourth, private party sellers and transactions at gun shows do not require FBI background checks in all states (Lang 2013). Finally, there may be temporal differences between when a background check is conducted and when a weapon is purchased. 10.7% of FBI background checks in 2017 were delayed by incomplete criminal records (FBI 2017a).

Table 5 presents state/month correlations between FBI background checks, LFP and the other candidate proxies at the state/month level for 2010-2017. The most important entry in the table is the correlation between FBI background checks and LFP at 0.41. This is larger than any of the suicide-based proxy correlations over the same period, as those range from 0.03-0.28. However, it also falls short of the STR correlation with LFP of 0.76 over the same time range.

FBI background checks correlate with STR at 0.80 and with OS at 0.80. Figure 5 illustrates the relationship between FBI background checks and STR, showing that they trend similarly with corresponding spikes and dips, but a difference in levels during the first four years of the sample seems to decrease during the second half of the sample.

Table 5 State-month correlations between background checks, LFP, and candidate firearm proxies

Proxy	ID	1	2	3	4	5	6	7	8	9
LFP/Pop.	1	1.00								
FBI Background Checks/Pop.	2	0.41*	1.00							
FSS	3	0.03	0.05	1.00						
FS/Pop.	4	0.17	0.04	0.84*	1.00					
S/Pop.	5	0.28*	0.00	−0.05	0.49*	1.00				
STR/Pop.	6	0.76*	0.80*	0.03	0.07	0.10	1.00			
OS/Pop.	7	0.47*	0.80*	−0.02	−0.04	−0.02	0.77*	1.00		
FFL/Pop.[‡]	8	0.73*	0.26	−0.13	0.12	0.27	0.42*	0.03	1.00	

*$p < 0.05$. [‡]:Data available from 2014-2017, with Sept., Oct. 2015 missing.

Fig. 5 FBI Background Checks vs. STR. Note. This figure shows the monthly trend for each FBI background checks and STR.
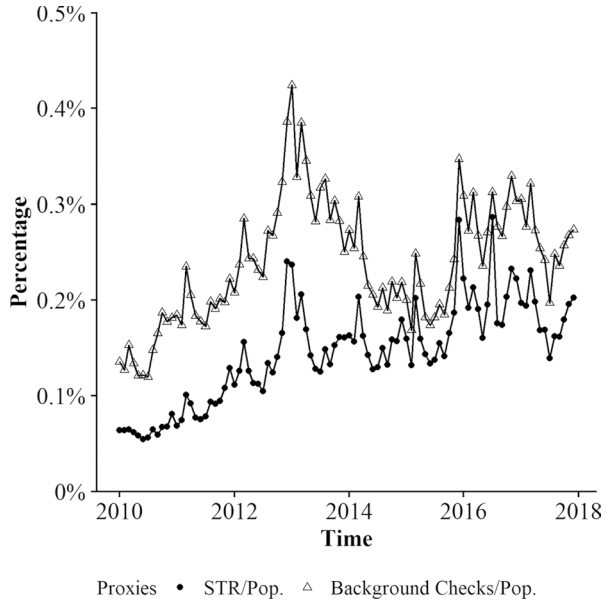
Overall, the state/month analysis shows that FBI background checks offer a better proxy for LFP than suicide-based proxies, but it is not as strong as other retail-based proxies.

5 LFP and proxy effects in homicide regressions

In this section we explore the role of proxy choice in a classic research design, the study of guns and homicide by Cook and Ludwig (2006; hereafter, "CL"). We estimate the association between LFP and homicide and then compare it with proxy estimates, holding data and methods constant.

CL analyzed annual National Center of Health Statistics (NCHS) and Uniform Crime Reports (UCR) data from 1980-1999 in the 200 most populous US counties. We focus on the model CL reported in Table 2, Column 2, in which homicide was regressed on lagged firearm prevalence, robberies, burglaries, year fixed effects and county fixed effects. The model is:

$$ln(Hom_{ct}) = \beta_0 + \beta_1 ln(Firearms_{ct-1}) + \beta_2 ln(Rob_{ct}) + \beta_3 ln(Burg_{ct}) + \alpha_c + \lambda_t + \epsilon_{ct} \tag{5}$$

Errors are population-weighted and clustered by county. No data source measures total firearm prevalence directly, so CL used FSS to proxy for TFP, finding that $\hat{\beta}_1 = 0.107$, with a 95% confidence interval of (0.034, 0.180).

We apply the CL research design with two key differences. First, we compare the LFP estimate to candidate proxies' estimates, rather than seeking to proxy for TFP.

Table 6 LFP and proxy effect estimates on homicide in county/month CL regressions

	Homicide					
	(1)	(2)	(3)	(4)	(5)	(6)
LFP	−0.461					
	(0.787)					
STR		−0.093				
		(0.268)				
Cumulative STR			−0.501			
			(1.289)			
Online Sales (OS)				0.042		
				(0.053)		
Cumulative OS					−0.066	
					(0.140)	
FSS						−0.264**
						(0.108)
Robbery	0.060	0.060	0.064	0.064	0.061	0.053
	(0.093)	(0.091)	(0.097)	(0.094)	(0.089)	(0.098)
Burglary	−0.047	−0.049	−0.036	−0.057	−0.042	−0.064
	(0.161)	(0.179)	(0.136)	(0.172)	(0.158)	(0.170)
Constant	5.359	1.300	5.302	0.639	0.849	0.878
	(8.698)	(1.578)	(12.509)	(1.121)	(1.212)	(1.04)
Time Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	756	756	756	756	756	756
R^2	0.549	0.549	0.549	0.550	0.549	0.552
Adjusted R^2	0.477	0.477	0.477	0.477	0.477	0.480

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

All variables are logged. Each proxy indicates the first lag. The sample is the 8 MA counties that are part of the top 171 largest US counties that reported UCR data continuously.

Second, we focus on monthly data for the 8 Massachusetts counties that were in the CL sample.[16] Our sample is more recent, has better temporal resolution, and measures LFP directly, whereas the original CL sample contained more counties and years.[17]

Table 6 shows the effect of LFP on homicide in the CL empirical framework, as well as estimates for the five best or most frequently used candidate proxies. The comparisons hold data and methods constant; only the measure or proxy for LFP changes across columns. The estimated effect of LFP on homicide is −0.461, with

---

[16] We have also estimated the CL model using FSS, OS and cumulative OS proxies in annual data among the 171 most populous counties with continuous UCR reporting from 2010-2017. The association between FSS and Homicide in this larger sample was 0.114 with a confidence interval of (0.007, 0.221), which is statistically significant and quantitatively similar to CL's finding.

[17] We add one to variables that contain zero values to avoid taking the log of zero. Four observations of FSS are undefined due to zero suicides, so we drop them from all regressions reported.

a 95% confidence interval of (−2.321, 1.400). The confidence interval for the estimated LFP effect on homicide is wide, so the sign of the effect is uncertain given considerable estimation error.

Table 6 also reports associations of homicide with five proxies for LFP: STR, Cumulative STR, OS, Cumulative OS, and FSS. Among these candidate proxies, the point estimate of Cumulative STR is −0.501, again most similar to the LFP effect, with a larger standard error of 1.289 leading to a 95% confidence interval of (−3.548, 2.546). Therefore, Cumulative STR would be an accurate but conservative proxy for LFP in this setting.

The results show that other candidate proxies yield overly precise estimates. The 95% confidence intervals for STR, OS, and Cumulative OS are all too narrow, at (−0.727, 0.540), (−0.084, 0.168), and (−0.396, 0.264), respectively. Therefore, these proxies' confidence intervals partially overlap with LFP's confidence interval. They are suboptimal given their overly narrow bounds but they accurately refrain from yielding a statistically significant association.

Surprisingly, the static FSS proxy produces a negative, significant association between LFP and homicide, with a 95% confidence interval of (−0.520, −0.007). County/month data is not ideal for FSS measurement, as granularity exacerbates the measure's volatility (Hayo et al. 2019; Cook and Ludwig 2019). Still, it is interesting that the proxy would lead to a Type I error, and also that its negative sign opposes CL's original significant, positive finding.

Table 7 reports estimated associations between the remaining proxies and LFP.[18] Five of these proxies accurately produce non-findings, but the Cumulative FFL proxy produces a positive, significant association with homicide, again due to underestimation of true parameter uncertainty. This result shows that use of an invalid proxy may produce either a positive or a negative spurious finding. These findings pertain to a single empirical setting, but they indicate that selection of a suboptimal proxy in a classic research design can underestimate parameter uncertainty and even cause a Type 1 error. They underscore the need for caution in proxy validation and selection. Ideally, we could systematically compile and publish valid firearm acquisition data in order to accurately estimate firearm policy effects on legal firearm acquisitions.

6 Discussion and implications

We introduce the first measure of Legal Firearm Prevalence. We introduce several new retail-based candidate proxies for LFP. We offer the first empirical evaluation of candidate proxies for LFP. We find that cumulative firearm acquisitions are the best proxy for LFP in every research design tested. Online sales of firearms can also proxy well for LFP, despite a small overall share of the market. Like prior work on TFP, we find that FSS is a good cross-sectional proxy and a poor

---

[18] The static version of FFL, the change in number of FFLs, is not included in the analysis because static FFL includes negative values and therefore cannot be logged.

Table 7 More proxy effect estimates on homicide in county/month CL regressions

	Homicide					
	(1)	(2)	(3)	(4)	(5)	(6)
Cumulative FSS	− 0.663					
	(0.902)					
Suicides (S)		0.048				
		(0.054)				
Cumulative S			0.245			
			(0.279)			
Firearm Suicides (FS)				− 0.045		
				(0.034)		
Cumulative FS					− 0.053	
					(0.269)	
Cumulative FFL						0.685***
						(0.077)
Robbery	0.052	0.060	0.053	0.055	0.058	0.055
	(0.091)	(0.092)	(0.091)	(0.094)	(0.091)	(0.082)
Burglary	− 0.056	− 0.054	− 0.056	− 0.061	− 0.053	− 0.007
	(0.174)	(0.178)	(0.175)	(0.167)	(0.175)	(0.218)
Constant	0.908	0.676	0.473	0.839	0.779	− 2.387**
	(0.957)	(1.024)	(0.840)	(1.004)	(0.975)	(1.374)
Time Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	756	756	756	756	756	360
R^2	0.550	0.550	0.550	0.550	0.549	0.556
Adjusted R^2	0.478	0.478	0.478	0.478	0.477	0.477

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

All variables are logged. Each proxy indicates the first lag. The sample is the 8 MA counties that are part of the top 171 largest US counties that reported UCR data continuously.

cross-temporal proxy for LFP. We showed that suboptimal proxies may lead to mistaken findings in applied research.

Next, we discuss implications for proxy selection, evaluation of firearm research, feasible policies to increase firearm sales data availability and limitations.

6.1 Implications for LFP proxy selection

This paper offers clear guidance for researchers who wish to study LFP as a predictor or outcome variable. Table 8 summarizes the results of proxy evaluations across research designs, using our preferred criteria for proxy validity: Partial R^2 greater than .5, a statistically significant parameter estimate, and positive association between proxy and LFP.

Table 8 Proxy summary across common designs

Proxy	Cross-section	State/Month Intertemporal	County/Year Panel	County/Year Panel	County/Month Panel	County/Month Panel
	Correlation	Correlation	(No Controls)	(with Controls)	Contemporaneous proxy	Lagged proxy
FSS	✓					
Cum. FS/ Cum. S	✓					
FS	✓					
Cumulative FS	✓	✓	✓	✓	✓	✓
S	✓	✓	✓			
Cumulative S	✓	✓	✓	✓	✓	✓
STR	✓	✓	✓	✓	✓	✓
Cumulative STR	✓	✓	✓	✓	✓	✓
OS	✓	✓	✓			
Cumulative OS	✓	✓	✓	✓	✓	✓
Change in FFL	✓		✓			
FFL	✓	✓				

Checkmark indicates the p-value of the correlation coefficient is less than .05 for cross-sectional and time-series designs. In panel designs, the checkmark indicates that the proxy exhibits a Partial R^2 greater than .5, a statistically significant parameter effect and a positive association with LFP.

One implication for selecting a proxy for LFP should be uncontroversial: If possible, choose legal firearm acquisitions as a proxy. It was an empirical question whether firearm sales would reflect the large served available market of legal firearms prevalence, but the data indicated STR as the best available proxy in every research design tested. Multiple states maintain firearm transaction registries; see, e.g., Sorenson and Berk (2001) or Berk (2021).

Most states do not register firearm transactions. Therefore, the finding that online firearm sales serve as a valid proxy for LFP in many research designs, despite having just a 2.7% overall market share, is quite promising. We anticipate that new sources of firearm sales may become available in the future. Possible sources include systematic collection of online firearm sales listings, large consumer expenditure panels, credit card transaction data or retailer parking capacity utilization. All of those data sources have been analyzed in other academic research, though not yet within the context of firearms. We recommend inclusion of firearm retail measures in future assessments of firearm data (e.g., NORC 2021).

We also found that FBI background checks correlate highly, at 0.41, with LFP. FBI background checks also correlate with STR and OS, each at 0.80, respectively, showing that background checks are a very good indicator of overall firearm sales. However, the analysis also illustrates limitations of the FBI background checks, since data are only available by state and month. Therefore, they cannot be used in more granular county-level, daily or weekly analyses.

Finally, the suicide-based proxy results provide several clear guidelines. FSS is only a valid proxy for LFP in cross-sectional research designs. Its selection could lead to mistaken conclusions in cross-temporal designs by overcontrolling for LFP. Surprisingly, despite the non-validation of FSS as an intertemporal proxy, the results show that accumulations of the two components of FSS – namely, Firearm Suicides and Suicides – both perform well in all research designs tested. It seems logical that if FSS contains valid but noisy signals about firearm prevalence, the accumulation of those signals may offer helpful information about local changes in firearm prevalence. However, we reach this conclusion with two strong caveats. First, the finding emerged *ex post*; we did not expect to find this *ex ante*. Second, we do not know any previous analysis that made the same point or found the same result, despite numerous papers that sought to study firearm prevalence in longitudinal designs. Therefore we would encourage further research on this topic to deepen our collective understanding.

6.2 Evaluation of published firearm research

FSS is the most common cross-sectional proxy for firearm prevalence, as validated by comparisons to survey measures of TFP. We found that FSS also correlates cross-sectionally with LFP. Therefore, cross-sectional analyses that use FSS to proxy for TFP could additionally be interpreted as applying to LFP.

We also found that static FSS is nearly uncorrelated with LFP over time, and FSS was not indicated as a valid proxy in any cross-temporal design tested. Further, the analysis in Section 5 found that FSS estimates were misleadingly precise

compared to LFP. Therefore, published results that used FSS in intertemporal designs may be unrepresentative of LFP.

6.3 Feasible policies to increase firearm sales data availability

Evaluations of firearm acquisition policies' effects on legal firearm acquisitions require a measure or proxy for LFP. LFP also may be relevant as a moderating variable between policy enactment and other potential policy-relevant outcomes, such as firearm suicides, assaults, defensive gun uses, mass shootings or other variables. A recent report found that "the current firearms data environment is disordered and highly segmented" (NORC 2021).

- The Federal Bureau of Investigation could publish background check data at more granular levels, such as county, city, zip code, week and date.
- States that collect firearm acquisition data, such as California and Massachusetts, could publish granular counts of firearm transactions.
- States, counties or cities that do not collect firearm acquisition data could collect and publish such data.
- Firearm retailers, retail chains or retailer associations could publish aggregate sales data by place and time.
- Digital platforms, advocacy groups or researchers could scrape, track and report online firearm sales.

We would advise caution in designing procedures for firearm acquisitions data collection, reporting and distribution. Firearm market participants do not always comply with restrictive policies (Balakrishna and Wilbur 2021), so we would recommend transparently safeguarding individual privacy. We would also advise that governments and retailers work to develop a coherent data reporting approach in order to maximize comparability of firearm acquisitions data across geography and time. We are confident that such concerns can be resolved, as they have already been considered and addressed in other sensitive contexts, such as the CDC's mortality data.

We speculate that political leadership and advocacy will be required before such actions become policy. Firearm data production and research has been controversial in the past. A Congressional amendment in 1996 stipulated that "none of the funds made available for injury prevention and control at the Centers for Disease Control and Prevention (CDC) may be used to advocate or promote gun control." The law effectively froze firearm research, to the later regret of the amendment's author (Inskeep 2015).

We believe that collecting and publishing privacy-compliant firearm acquisition data would be a helpful step in evaluating firearm acquisition and use policies. It could also help reveal the mechanisms between policy changes and criminal or health outcomes. As an example, it could show how concealed-carry, open-carry or stand-your-ground policies affect legal firearm acquisitions, and then how changes

in legal firearm prevalence affect crime, violence or health. However, such analyses may remain incomplete without IFP or TFP data.

6.4 Limitations

The current article has several important limitations. We were only able to collect LFP and STR data for Massachusetts. Massachusetts has some of the strictest firearm laws and lowest firearm ownership rates in the nation, so it remains to be seen whether the results will generalize to other jurisdictions.

We do not have measures of IFP, so we are not able to make empirical statements about TFP or IFP. We expect that IFP will be more strongly related to illegal firearm uses than LFP. We also suspect that LFP and IFP may be related, as increasing LFP may increase supply in illegal firearm markets.

A smaller concern is that, although LFP could have decreased during the sample period, it rose nearly uniformly in every county throughout the observation window. Cumulative proxy variables may perform worse if firearm prevalence sometimes decreases. However, firearms are durable goods and we do not know of any periods of sustained firearm divestitures in recent U.S. history, therefore we are not certain how important this limitation may be. It certainly could become important if, for example, jurisdictions bought back or outlawed particular types of firearms. However, if such policies were enacted, public records might provide firearm divestitures data which could then be used to adjust LFP measures or proxies.

6.5 Conclusion

Firearm acquisition and usage policy evaluations require reliable, systematically collected, longitudinally valid proxies for legal firearm prevalence. We hope the present paper can advance firearm research by evaluating proxies for LFP. We hope that collaborative efforts between policy makers, researchers and data providers will enhance scientific knowledge and provide the evidence needed to help evaluate and inform firearm policy.

Appendix A: Backfilling license data

Massachusetts' firearm license records date back to January 1, 2006. Therefore, the data enable direct counts of licensees after January 1, 2012, as licenses remain valid for 6 years after issuance. However, the 2010 and 2011 data undercount the true numbers of firearm licensees, as they exclude valid licenses issued in 2004 and 2005. This manifests in the data as a misleadingly steep upward trend in 2010-11, as depicted in panel (a) in Fig. 6.

To address left censoring in 2010 and 2011, we first checked how often firearms licenses are renewed, expecting a high renewal rate. The data confirm that
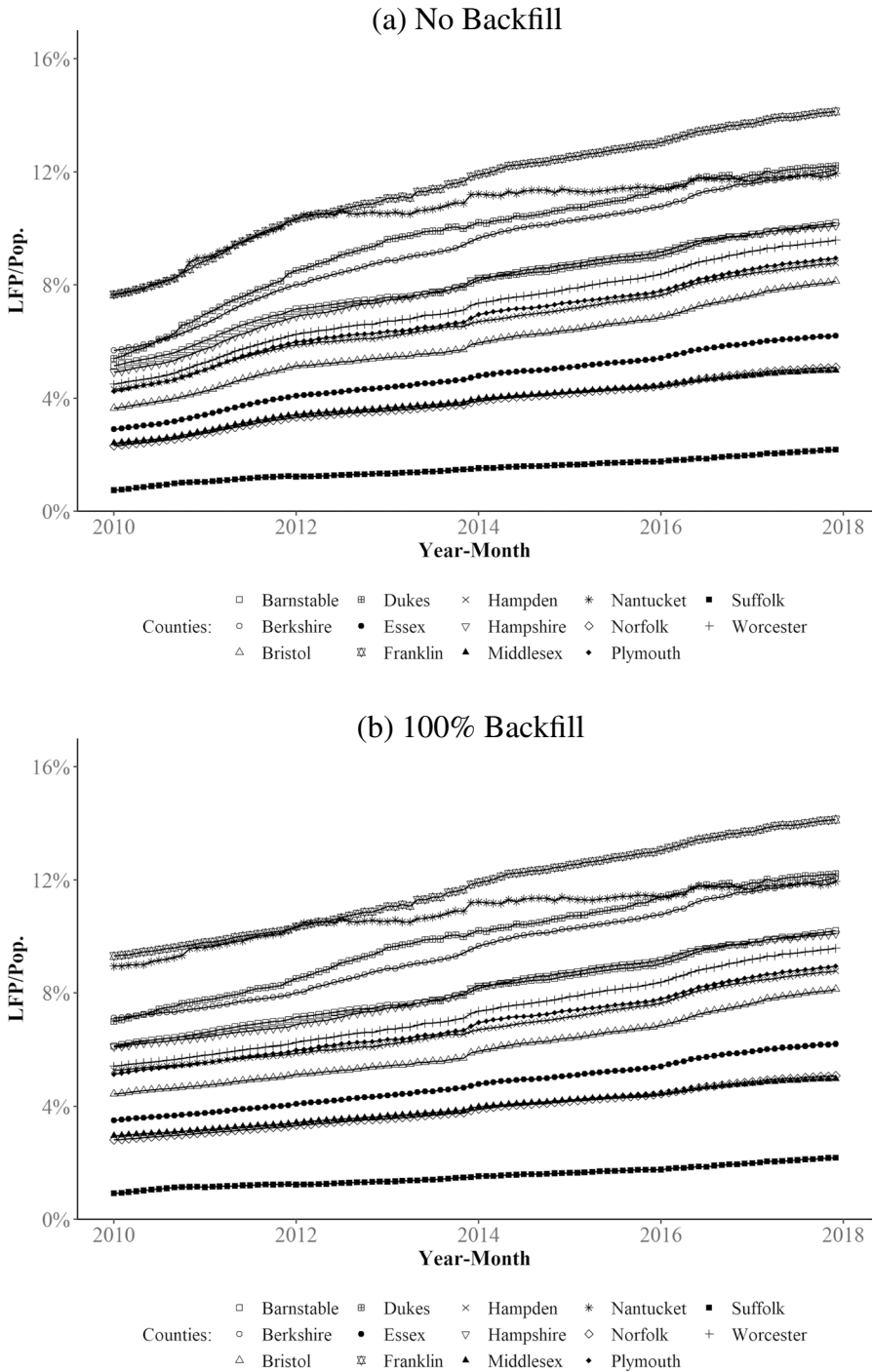
Fig. 6 100% Backfill vs. No Backfill: LFP trend by county. Note. Panel (a) shows LFP over time for each MA county using the raw data, while Panel (b) shows that using backfilled data.

expectation: among all license renewals observed after 2012, 94% of licensees are observed to hold at least one valid license between 2006 and 2012. We therefore resolve the left-censoring issue by "backfilling" all resident license renewals observed in 2010 and 2011. For example, a licensee who is issued a license renewal in February 2011 is also counted as an active licensee from January 2010 until January 2011. After backfilling, the LFP trend looks much smoother as shown in panel (b) in Fig. 6. We have replicated the analysis in section 4 without any backfilling, and with stochastic backfilling of 94% of licenses. The empirical results are relatively insensitive to which method of backfilling is used.

Appendix B: Stability of cross-sectional correlations

Here we examine the stability of cross-sectional correlations across successive time periods. The goal is to illustrate why FSS can be both a good cross-sectional proxy for LFP and a poor cross-temporal proxy for LFP.

We report the following exercise: Suppose that we only had two years of data available; how much would the cross-sectional correlations depend on which two years we analyze? We partition the sample into four distinct two-year periods and calculate cross-sectional correlations within each.

Table 9 shows that STR is not only the best cross-sectional proxy for LFP overall, it is also the most stable across subsamples. The four cross-sectional correlations between STR and LFP range from .82 to .86, a range that contains the overall eight-year correlation of .86. OS is less stable, with correlations ranging from .46 to .64. FFL can only be measured in two partitions, correlating with LFP at .55 and .59.

Firearm Suicides is the most stable suicide-based proxy, with two-year correlations ranging from .39 to .58. Correlations between LFP, FSS and Suicides are substantially more volatile. FSS correlates with LFP at .26 in 2010-11 and at .66 in 2016-17. Remarkably, all four two-year correlations between FSS and LFP are smaller than the eight-year correlation of .74, and only one of the four is statistically significant. The correlation between Suicides and LFP also ranges widely from .36 to .79.

Figure 7 graphs LFP and FSS for the fourteen counties within each two-year partition. The relationships are influenced by outliers, including three observations of zero firearm suicides in one county, and one observation of FSS equal to one. FSS appears to be particularly prone to outliers (e.g., 1) or indeterminacy (i.e., 0/0) when measured in more granular samples than the other candidate proxies, confirming limitations on the geographic and temporal resolutions in which it may be applied.

Table 9 Biennial cross-sectional correlations between LFP and candidate firearm proxies

Proxy	ID	1	2	3	4	5	6	7
Panel 1: 2010-2011								
LFP/Pop.	**1**	1.00						
FSS	**2**	0.26	1.00					
FS/Pop.	**3**	0.39	0.95*	1.00				
S/Pop.	**4**	0.48	0.07	0.36	1.00			
STR/Pop.	**5**	0.86*	0.29	0.41	0.42	1.00		
OS/Pop.	**6**	0.46	0.54*	0.64*	0.29	0.59*	1.00	
FFL/Pop.‡	**7**	NA	NA	NA	NA	NA	NA	1.00
Panel 2: 2012-2013								
Proxy	ID	1	2	3	4	5	6	7
LFP/Pop.	**1**	1.00						
FSS	**2**	0.38	1.00					
FS/Pop.	**3**	0.49	0.96*	1.00				
S/Pop.	**4**	0.43	0.27	0.49	1.00			
STR/Pop.	**5**	0.82*	0.20	0.33	0.50	1.00		
OS/Pop.	**6**	0.64*	−0.43	−0.32	0.14	0.60*	1.00	
FFL/Pop.‡	**7**	NA	NA	NA	NA	NA	NA	1.00
Panel 3: 2014-2015								
Proxy	ID	1	2	3	4	5	6	7
LFP/Pop.	**1**	1.00						
FSS	**2**	0.38	1.00					
FS/Pop.	**3**	0.58*	0.94*	1.00				
S/Pop.	**4**	0.79*	0.48	0.74*	1.00			
STR/Pop.	**5**	0.82*	0.37	0.47	0.58*	1.00		
OS/Pop.	**6**	0.64*	−0.09	0.11	0.47	0.62*	1.00	
FFL/Pop.‡	**7**	0.55*	0.52	0.55*	0.43	0.71*	0.50	1.00
Panel 4: 2016-2017								
Proxy	ID	1	2	3	4	5	6	7
LFP/Pop.	**1**	1.00						
FSS	**2**	0.66*	1.00					
FS/Pop.	**3**	0.41	0.83*	1.00				
S/Pop.	**4**	0.36	0.76*	0.99*	1.00			
STR/Pop.	**5**	0.85*	0.39	0.05	−0.01	1.00		
OS/Pop.	**6**	0.59*	−0.05	−0.20	−0.19	0.65*	1.00	
FFL/Pop.‡	**7**	0.59*	0.17	−0.26	−0.33	0.79*	0.36	1.00

*$p < 0.05$. 14 observations (1 per county) are used to calculate each correlation. ‡:Data available from 2014-2017, with Sept., Oct. 2015 missing.
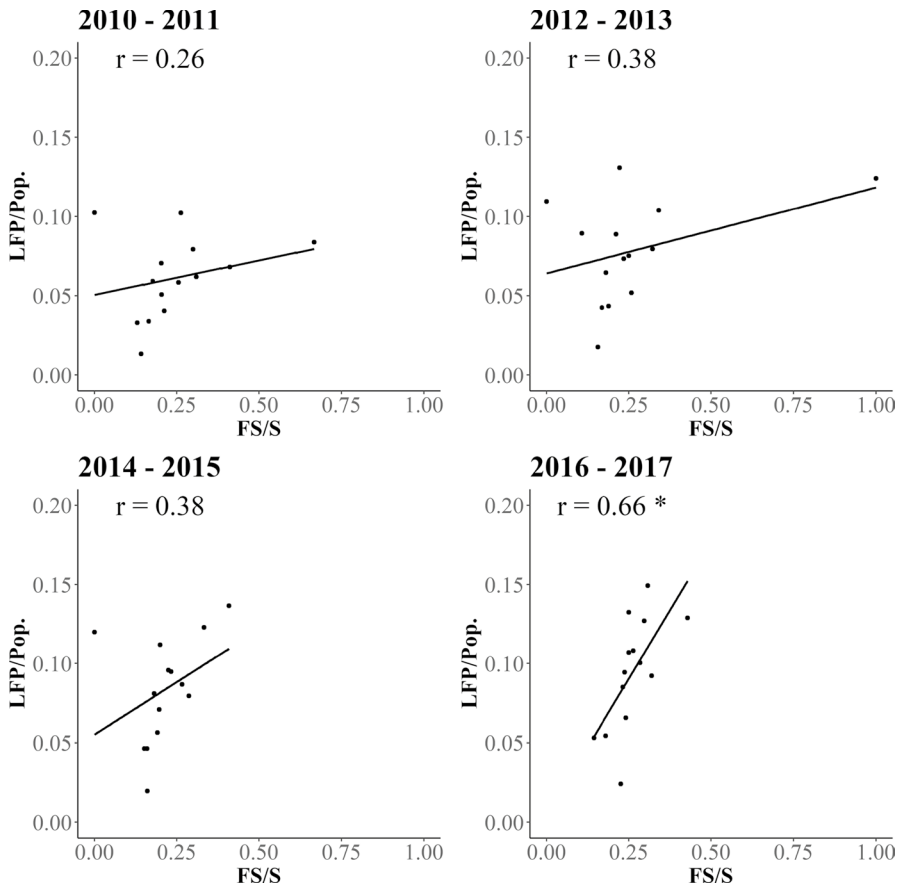
Fig. 7 Scatterplots of FSS and LFP - Biennial Comparisons. Note. Each point represents a county. r is the cross-sectional Pearson correlation score between FSS and LFP/Pop. *$p < 0.05$.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Andrés, A. R., & Hempstead, K. (2011). Gun control and suicide: The impact of state firearm regulations in the united states, 1995–2004. *Health Policy*, *101*(1), 95–103.

Azrael, D., Cook, P. J., & Miller, M. (2004). State and local prevalence of firearms ownership measurement, structure, and trends. *Journal of Quantitative Criminology*, *20*(1), 43–62.

Balakrishna, M., & Wilbur, K. C. (2021). Do firearm markets comply with firearm restrictions? how the massachusetts assault weapons ban enforcement notice changed firearm sales. Journal of Empirical Legal Studies.

Bangalore, S., & Messerli, F. H. (2013). Gun ownership and firearm-related deaths. *The American Journal of Medicine*, *126*(10), 873–876.

Berk, R. A. (2021). Post-model-selection Statistical inference with interrupted time series designs: An evaluation of an assault weapons ban in california. 2105.10624 accessed Oct 2021.

Bordua, D. J. (1986). Firearms ownership and violent crime: A comparison of illinois counties. The Social Ecology of Crime, 156–188 Springer.

Briggs, J. T., & Tabarrok, A. (2014). Firearms and suicides in us states. *International Review of Law and Economics*, *37*, 180–188.

Castillo-Carniglia, A., Kagawa, R. M. C., Webster, D. W., Vernick, J. S., Cerda, M., & Wintemute, G. J. (2018). Comprehensive background check policy and frearm background checks in three us states. *Injury Prevention*, *24*, 454–459.

Cook, P. J. (1979). The effect of gun availability on robbery and robbery murder: a cross-section study of fifty cities. *Policy Studies Review Annual*, *3*, 743–781.

Cook, P. J., & Ludwig, J. (2006). The social costs of gun ownership. *Journal of Public Economics*, *90*(1-2), 379–391.

Cook, P. J., & Ludwig, J. (2019). The social costs of gun ownership: a reply to hayo, neumeier, and westphal. *Empirical Economics*, *56*(1), 13–22.

Depetris-Chauvin, E. (2015). Fear of obama: an empirical study of the demand for guns and the us 2008 presidential election. *Journal of Public Economics*, *130*, 66–79.

Donohue, J. J., Aneja, A., & Weber, K. D. (2019). Right-to-carry laws and violent crime: a comprehensive assessment using panel data and a state-level synthetic control analysis. *Journal of Empirical Legal Studies*, *16*(2), 198–247.

Duggan, M. (2001). More guns, more crime. *Journal of Political Economy*, *109*(5), 1086–1114.

FBI. (2017a). 2017 nics operations report. https://www.fbi.gov/file-repository/2017-nics-operations-report.pdf/view, accessed: 2020-11-10.

FBI. (2017b). Total nics firearm background checks by state. https://www.fbi.gov/services/cjis/nics, accessed: 2020-11-10.

Fisher, J. C. (1976). Homicide in detroit the role of firearms. *Criminology*, *14*(3), 387–400.

Gilson, D. (2021). New data shows that the nra's trump bump has evaporated. Mother Jones https://www.motherjones.com/politics/2021/01/national-rifle-association-membership-magazines/, accessed Oct 2021.

Haas, S. M., Jarvis, J. P., Jefferis, E., & Turley, E. (2007). Gun availability and crime in west virginia: an examination of nibrs data. *Justice Research and Policy*, *9*(2), 139–159.

Hayo, B., Neumeier, F., & Westphal, C. (2019). The social costs of gun ownership revisited. *Empirical Economics*, *56*(1), 1–12.

Ingraham, C. (2021). Nobody knows how many members the nra has, but its tax returns offer some clues. Washington Post. https://www.washingtonpost.com/news/wonk/wp/2018/02/26/nobody-knows-how-many-members-the-nra-has-but-its-tax-returns-offer-some-clues/. Accessed Oct 2021.

Inskeep, S. (2015). Ex-rep. dickey regrets restrictive law on gun violence research. https://www.npr.org/2015/10/09/447098666/ex-rep-dickey-regrets-restrictive-law-on-gun-violence-research, accessed Oct 2021.

Khalil, U. (2017). Do more guns lead to more crime? Understanding the role of illegal firearms, *Journal of Economic Behavior & Organization, 133*,342-361. https://doi.org/10.1016/j.jebo.2016.11.010

Kleck, G. (2004). Measures of gun ownership levels for macro-level crime and violence research. *Journal of Research in Crime and Delinquency*, *41* (1), 3–36.

Kleck, G. (2015). The impact of gun ownership rates on crime rates: a methodological review of the evidence. *Journal of Criminal Justice*, *43* (1), 40–48.

Kovandzic, T., Schaffer, M. E., & Kleck, G. (2011). Gun prevalence, homicide rates and causality: A gmm approach to endogeneity bias. The Sage Handbook of Criminological Research Methods.

Kovandzic, T., Schaffer, M. E., & Kleck, G. (2013). Estimating the causal effect of gun prevalence on homicide rates: a local average treatment effect approach. *Journal of Quantitative Criminology*, *29*(4), 477–541.

Kronmal, R. A. (1993). Spurious correlation and the fallacy of the ratio standard revisited. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *156*(3), 379–392.

Krug, A. S. (1967). A statistical study of the relationship between firearms licensing laws and crime rates. Congressional Record 7–25.

Lang, M. (2013). Firearm background checks and suicide. *The Economic Journal*, *123*(573), 1085–1099.

McDowall, D. (1991). Firearm availability and homicide rates in detroit, 1951–1986. *Social Forces*, *69*(4), 1085–1101.

Moody, C. E., & Marvell, T. B. (2005). Guns and crime. Southern Economic Journal 720-736.

Moody, C. E., et al. (2010). Firearms and homicide. Handbook on the Economics of Crime, 432–451.

NORC. (2021). The state of firearms data in 2019. https://www.norc.org/PDFs/Firearm%20Data%20Infrastructure20Expert%20Panel/State%20of%20Firearms%20Research%202019.pdf, Accessed Oct 2021.

Powell, K. E., Jacklin, B. C., Nelson, D. E., & Bland, S. (1998). State estimates of household exposure to firearms, loaded firearms, and handguns, 1991 through 1995. *American Journal of Public Health*, *88*(6), 969–972.

Research, P. (2021). Amid a series of mass shootings in the u.s., gun policy remains deeply divisive. https://www.pewresearch.org/politics/2021/04/20/amid-a-series-of-mass-shootings-in-the-u-s-gun-policy-remains-deeply-divisive/. Accessed Oct 2021.

Schell, T. L., Peterson, S., Vegetabile, B. G., Scherling, A., Smart, R., & Morral, A. R. (2020). State-level estimates of household firearm ownership. Technical report, RAND Corporation.

Siegel, M., Ross, C. S., & King, C. III (2013). The relationship between gun ownership and firearm homicide rates in the united states, 1981–2010. *American Journal of Public Health*, *103*(11), 2098–2105.

Siegel, M., Negussie, Y., Vanture, S., Pleskunas, J., Ross, C. S., & King, C. III (2014a). The relationship between gun ownership and stranger and nonstranger firearm homicide rates in the united states, 1981–2010. *American Journal of Public Health*, *104*(10), 1912–1919.

Siegel, M., Ross, C. S., & King, C. (2014b). A new proxy measure for state-level gun ownership in studies of firearm injury prevention. *Injury Prevention*, *20*(3), 204–207.

Sloan, J. H., Kellermann, A. L., Reay, D. T., Ferris, J. A., Koepsell, T., Rivara, F. P., Rice, C., Gray, L., & LoGerfo, J. (1988). Handgun regulations, crime, assaults, and homicide. *New England Journal of Medicine*, *319* (19), 1256–1262.

Small Arms Survey. (2018). Small arms survey reveals: More than one billion firearms in the world. http://www.smallarmssurvey.org/leadmin/docs/Weapons_and_Markets/Tools/Firearms_holdings/SAS-Press-release-global-rearms-holdings.pdf. Accessed Oct 2021.

Smart, R., Morral, A. R., Smucker, S., Cherney, S., Schell, T. L., Peterson, S., Ahluwalia, S. C., Cefalu, M., Xenakis, L., Ramchand, R., & Gresenz, C. R. (2020). The science of gun policy: a critical synthesis of research evidence on the effects of gun policies in the united states, second edition. Technical report, RAND Corporation.

Sorenson, S. B., & Berk, R. A. (2001). Handgun sales, beer sales, and youth homicide, california, 1972–1993. *Journal of Public Health Policy*, *22*(2), 182–197.

Stolzenberg, L., & D'alessio, S. J. (2000). Gun availability and violent crime: New evidence from the national incident-based reporting system. *Social Forces*, *78*(4), 1461–1482.

Vitt, D. C., McQuoid, A. F., Moore, C., & Sawyer, S. (2018). Trigger warning: the causal impact of gun ownership on suicide. *Applied Economics*, *50* (53), 5747–5765.