# Navigating Potential Pitfalls in Difference-in-Differences Designs: Reconciling Conflicting Findings on Mass Shootings' Effect on Electoral Outcomes

HANS J. G. HASSELL *Florida State University*
JOHN B. HOLBEIN *University of Virginia*

*Recent work on the electoral effects of gun violence in the United States relying on the difference-in-differences design has produced starkly conflicting findings that range from null effects to substantively large effects. At the same time, as the difference-in-difference design on which this research has relied has exploded in popularity, scholars have documented several methodological issues this design faces—including potential violations of parallel trends and unaccounted for treatment effect heterogeneity. Sadly, these pitfalls (and their solutions) have not been fully explored in political science. In this paper, we apply these advancements to the unresolved debate on the effects of gun violence on electoral outcomes in the United States. We show that studies that find a large positive effect of gun violence on Democratic vote share are a product of a failure to properly specify difference-in-differences models when their underlying assumptions are unlikely to hold. Once these biases are corrected, shootings show little evidence of sparking large electoral change. Our work clarifies an important unresolved debate and provides a road-map for the many scholars currently employing difference-in-differences designs.*

Word Count: 14,216

Dr. Hassell is Associate Professor of Political Science and the Director of the Institute of Politics at Florida State University, 32306 FL (hans.hassell@fsu.edu)

Dr. Holbein is Assistant Professor of Public Policy, Politics, and Education at the Frank Batten School of Leadership and Public Policy at the University of Virginia, 22902. Corresponding Author (holbein@virginia.edu).

We'd like to thank Laura Garcia-Montoya, Ana Arjona, Matthew Lacombe, and Hasin Yousaf for making their replication data and code available. We thank Matt Baldwin, Charles Crabtree, Bernard Fraga, Andrew Goodman-Bacon, Zoltan Hajnal, Nazita Lajevardi, Gareth Nellis, Zayne Sember, Neil Visalvanich, Bertrand Wilden, and Yiqing Xu for their feedback.

**This is a manuscript submitted for review.**

Gun violence in the United States has a devastating impact on communities, families, and individuals (e.g., Rossin-Slater et al. 2020). Yet, despite repeated tragedies and public support for policies that would reduce gun violence, the policy response has been, at best, tepid. This presents an unsolved puzzle. Why do an abundance of salient mass shootings and a supportive public fail to result in meaningful policy change to address gun violence? To solve this puzzle, scholars have looked to see whether mass shootings change electoral incentives and whether a lack of policy changes occurs *in spite of* or (perhaps) *because of* a lack of electoral pressure. In estimating the effects of these tragic shootings on elections, scholars have relied on panel data and difference-in-differences designs that exploit variation in the timing and location of shootings. Yet, despite using the same data sources, previous work has reached starkly different conclusions, with some work finding that mass shootings have strong electoral effects (Garcia-Montoya et al. 2022; Yousaf 2021) and others finding null effects (Hassell et al. 2020).

In this paper, we argue that these conflicting findings are explained by the failure of some work to account for violations of the parallel trends assumption in their modeling decisions. Core assumptions to difference-in-difference approach, particularly the parallel trends assumption, are essential, non-negotiable requirements for causal inferences.[1] Simply, previous work documenting large effects of gun violence on electoral outcomes has done so erroneously because shootings are more likely to happen in areas already trending towards more pro-Democratic voting patterns long before shootings happened, whereas areas in the control group—i.e. areas where shootings haven't occurred—were naturally trending more Republican. This is likely because mass shootings disproportionately occur in growing populations and have been increasing over time (U.S. Government Accountability Office

[1]Modeling decisions are just one component of the many researcher degrees of freedom—the many seemingly small choices in designing, collecting, analyzing, and reporting results—and that can affect the conclusions drawn from difference-in-differences designs. As such, Gelman and Loken (2013) liken research to a "garden of forking paths." Although the parallel trends assumption is the primary driver in the differences in conclusions among work on the effects of mass shootings, as we outline below other researcher driven choices can also have effects on the conclusions reached. The effects of those choices (including how to measure the outcome, how to specify the treatment, the functional form of time trends, and how (or at what level) to adjust standard errors) can also have meaningful effects.



APSR Submission Template

2020; Musu-Gillette et al. 2018), combined with a realignment in American politics in which these same more populated areas have become consistently more Democratic over time (DeSilver 2016). As a result, TWFE designs suggest mass shootings have a large effect on vote share in presidential elections up to 20 years *before* these shootings happened, suggesting critical identification issues with a TWFE estimator in this particular case.

Models that account for these violations of parallel trends provide little to no evidence mass shootings cause large and meaningful electoral change in the United States and compelling evidence consistent with a null effect.[2] This result is consistent no matter whether we look at all school shootings, just "rampage-style" shootings, or mass shootings more generally.[3]

In addition to having value in its own right, resolving the stark differences in published findings on the effect of mass shootings on electoral outcomes also provides an useful case to illustrate the importance of properly navigating potential pitfalls in difference-in-differences designs. In recent years, the difference-in-differences design has proliferated as a means of elucidating causal effects, partially because of its intuitive simplicity and relatively modest data requirements coinciding with a broader interest in causal inference and "credible" estimates (Angrist and Pischke 2010).[4,5] Rapid growth in

[2]Though some corrected models are unable to rule out a much smaller effect, most of these estimates are not statistically significant and negative effects often show up across the small, but reasonable, changes to model specification well within reasonable researcher degrees of freedom. Moreover, sensitivity analyses that embrace the uncertainty around exact departures from parallel trends show that the results are *highly* sensitive to even *minimal* reasonable departures from parallel trends. Hence, the preponderance of evidence suggests that a large effect is implausible and smaller positive effects debatable at best.

[3]While claiming "[their] findings hold when [they] replicate [Hassell et al.'s 2020] models" (17) which include county-specific time-trends, Garcia and colleagues (2022) do not test models with county-specific time-trends in any of their models in their manuscript, appendix, or replication materials. Although not claiming he does, Yousaf (2021) also does not include county-specific time trends. None of the three prior papers apply more recent advances for addressing potential violations of the parallel-trends assumption.

[4]The data and design demands for executing a difference-in-differences are light, requiring no instrument that satisfies the exclusion restriction (as does instrumental variable analysis) nor precise cutoffs (as do regression discontinuity designs).

[5]In 2022, there were over 17,000 new papers (and over 100 in political science) employing or discussing this



APSR Submission Template

the use of difference-in-differences has prompted a growing methodological literature covering the potential and pitfalls of this design (e.g., Roth et al. 2022; Kahn-Lang and Lang 2020; De Chaisemartin and d'Haultfoeuille 2020).

Thus, while we are most interested in answering the question of whether mass shootings affect party vote shares in the U.S., another purpose of this paper is to help narrow the gap between theory and application in the use of the difference-in-differences design in political science. We do so by 1.) outlining general cautions related to potential biases that arise from A.) violations of parallel trends and B.) treatment effect heterogeneity that researchers should take when implementing difference-in-differences designs flowing from recommendations in other disciplines (Roth et al. 2022; Kahn-Lang and Lang 2020); 2) highlighting the importance of researcher degrees of freedom related to specifying difference-in-differences models such as how one codes the treatment, whether to include time trends at all and, if so, the functional form of those time trends, and how (or at what level) to adjust standard errors; and 3) implementing them in an applied example. The steps we outline flow from those recommended in other disciplines (Roth et al. 2022; Kahn-Lang and Lang 2020). Our work collates recent advances in the difference-in-differences literature in a single applied example.

Given our strong interest in a particular substantive question, the core of this exploration revolves around diagnosing and dealing with potential parallel trends assumption violations that arise in questions of the causal effects of gun violence on electoral outcomes. However, we also address other potential issues; for instance, issues that arise with heterogeneous treatment effects. Thus, we provide a synthesis of the potential issues that may arise in difference-in-differences designs and the tests one can use to check for the robustness of their research findings that use the difference-in-differences design, particularly in the face of potential parallel-trends assumption violations. We (briefly) outline the problems that necessitate these checks, describe the logic of these checks, lay out how these checks are conducted, and point scholars to the statistical tools and resources that have been developed to employ these checks in practice. This exercise provides an applied example that researchers can use as a guide in their application of the difference-in-differences design.

Our work helps provide an answer to an important unresolved debate, shedding light on the political

method across the sciences in a given year (and this number is growing as shown in the online appendix).



APSR Submission Template

economy of gun violence in the United States, and contributing to our understanding of what events spark electoral accountability as well as providing some guidance in navigating the promising, but also perilous, difference-in-differences design.

## 1. DIFFERENCE-IN-DIFFERENCES AND THE TWO-WAY FIXED EFFECT ESTIMATOR

Before jumping into our specific case and how the potential pitfalls in the difference-in-difference design explain the divergence in findings, we think it important to explain the predominant approach scholars use and why those pitfalls exist. Difference-in-differences designs routinely rely on what has been termed the *two-way fixed effects estimator* (i.e. the TWFE). With this estimator, the outcome of interest is regressed on group and period fixed effects, along with the treatment status one desires to estimate. The TWFE approach, using time and unit (often a geographic level) fixed effects, controls for factors remaining constant within years (e.g., nationwide economic conditions), given that the treatment is not uniformly given in a single instance in time, and for factors varying across spaces (e.g., stable local area culture) that might impact the effect of the treatment implemented. In the early years of difference-in-difference designs, these identification strategies were used in the context of largely exogenous policy interventions, where treatment was implemented in a single instance in time—uniform across all treated units. This design constitutes a two-group (i.e. treated and not treated) and two-period (pre and post) design. As De Chaisemartin and D'Haultfoeuille (2022, 3) note "in the two-groups and two-periods design..., [the difference-in-differences estimator] is equal to the treatment coefficient in a TWFE regression with group and period fixed effects."

Importantly, this design rests on the so called parallel trends assumption, which requires that in the absence of the treatment, both the treated and the untreated group would have experienced the same outcome evolution.[6] Because of the fundamental problem of causal inference, we do not observe the counter-factual worlds in which the treated and untreated groups are exposed to the opposite

[6]As Roth et al. (2022, 2) put it, in the canonical example "the key identifying assumption is that the average outcome among the treated and comparison populations would have followed 'parallel trends' in the absence of treatment." They also layer a related assumption which requires "that the treatment has no causal effect before its implementation (no anticipation)."



APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

condition. Hence, the parallel trends is (in a way) not fully testable. However, De Chaisemartin and D'Haultfoeuille (2022, 2) call the parallel trends assumption "partly testable" as researchers can "compar[e] the outcome trends of groups [the treatment groups] and [the control group], before [the treatment group] received the treatment."[7] As an extension of the two-groups and two-periods design, scholars have also estimated "TWFE regressions in more complicated designs with many groups and periods, variation in treatment timing, treatments switching on and off, and/or non-binary treatments" (De Chaisemartin and D'Haultfoeuille 2022, 3). Recent research has shown that for these more sophisticated designs, TWFE estimators are unbiased if both the parallel trends assumption holds and "the treatment effect [is] constant, between groups and over time" (De Chaisemartin and D'Haultfoeuille 2022, 3). As we show in more detail below, the failure to account for potential violations of the parallel-trends assumptions (and treatment effect heterogeneity to a lesser extent) ultimately effects the conclusions drawn about the effect of mass shootings on election outcomes.

## 2. DIFFERENCES IN FINDINGS ON THE ELECTORAL EFFECTS OF MASS SHOOTINGS

In an article published at the *American Political Science Review* (APSR), Hassell, Holbein, and Baldwin (2020; hereafter HHB) estimate the effect of school shootings (of various sizes) on voter turnout and election outcomes at the federal, state, and local level.[8] HHB find that school shootings—regardless of

[7]Marcus and Sant'Anna (2021) note that the answer of whether pretreatment tests are illuminative of the parallel trends assumption "depends on the chosen [parallel trends assumption" one makes. Marcus and Sant'Anna (2021) discuss parallel trends such as "parallel trends assumption across all time periods and all groups", the "parallel trends assumption based on 'never treated' units", and the "parallel trends assumption based on 'not-yet-treated' units." Each of these vary in terms of how informative tests of pre-treatment balance are. (See Marcus and Sant'Anna (2021, 241-245) for their full discussion of each of these variants.) For other discussions of the usefulness of pre-trends tests see Kahn-Lang and Lang (2020); Bilinski and Hatfield (2018) and Roth (2022) and for explorations of potential relaxations of the parallel trends assumption see Manski and Pepper (2018); Rambachan and Roth (2021) and Freyaldenhoven et al. (2019). (We return to relaxations of parallel trends in our empirical examinations below.)

[8]HHB also use regression discontinuity in time to assess effects of shootings on voter registration (also a null result).

the number of victims—have precisely-estimated null effects on vote shares. In contrast, in another recently published article in *APSR*, Garcia-Montoya, Arjona, and Lacombe (2022; hereafter GMAL) seek to isolate the effect of "rampage-style" school shootings and find a sizable effect of around 5 percentage points on Democratic vote share in presidential elections in the local communities in which they occur.[9] Moreover, some of the plausible alternatives to the specifications GMAL run suggest an effect as large as ≈ 8.7 percentage points in the election after a mass shooting occurs and some event-study models using a two-way fixed effect (TWFE) specification suggest an effect as large as 13 percentage points a full 28 years after a mass shooting occurs. Between these two estimates, in a article published in the *Journal of the European Economic Association*, Yousaf (2021) argues that all mass shootings—not restricted to those occurring at schools—decrease Republican vote share in presidential elections by 2-6 percentage points (Yousaf 2021).[10]

Ultimately, the conclusions of GMAL and Yousaf (albeit less so) stand in contrast to HHB's—with the former studies suggesting that mass shootings have statistically detectable meaningful effects in the local communities in which they occur, despite relying on the same outcome data (vote share recorded in Dave Leip's Atlas of U.S. Elections).[11] Our work here shows that results suggesting large

[9]GMAL emphasize the topline 5 percentage point effect in their abstract and throughout various summary points in their manuscript. Depending on the sample one uses—be it the full pool of observations or only those for which covariates are available—their naive TWFE are 5.5 and 4.5 percentage points respectively ($p < 0.001$ in both cases).

[10]Unlike HHB, neither GMAL nor Yousaf examine midterms or state or local races.

[11]We think important to note that all three papers have some common findings. GMAL, Yousaf, and HHB all find no effects of mass shootings on voter turnout (as we show in the Online Appendix (see Figure S10) and turnout does not appear to have the trend differences that plague Democratic vote share (Note: This overall finding is also corroborated by a recent working paper by Marsh (2022, 21), who finds that changes in turnout after mass shootings are "not statistically distinguishable from zero." While Marsh does provide some evidence that mass shootings close to an election have a slight positive effect on turnout (see Marsh 2022, Figure 1), HHB show that the effects of school shootings close to an election on turnout are highly sensitive to model specification (see HHB, Figure A7) a pattern also somewhat evident in Marsh's models (see Marsh, Table E2 and E5)). Importantly, then, given the lack of any substantive effect on turnout, any increase in Democratic vote share should come from persuasion, rather than mobilization, unless gun violence simultaneously demobilizes Republicans and mobilizes



APSR Submission Template

effects of mass shootings are not robust because they fail to fully account for violations of the critical parallel-trends assumption. We also illustrate how exploring heterogeneity in treatment effects can illuminate the question of how mass shootings shape elections (if at all).

## 3. DATA

In efforts to resolve discrepancies between previous findings on the electoral effects of gun violence and provide a guide for navigating the pitfalls of difference-in-differences designs, we begin by examining the key differences between all the previous work examining the electoral effects of gun violence.[12] All prior work uses a common dataset—Dave Leip's Atlas of U.S. Elections—to document electoral returns.

In estimating the difference-in-difference design, one source of researcher degrees of freedom is what counts as treatment. In our applied case, each study uses slightly different shootings as their treatment. We outline coding differences in the Online Appendix (see Table S1 in the Online Appendix for a summary of all the differences between the studies in data and methodological choices).[13] However, despite what some have previously claimed (GMAL 2022, 821-823), differences in data choices and

Democrats *at the exact same rates*, which is highly unlikely. However, any persuasive effect would also likely show up in attitudinal shifts and previous research on the attitudinal effects of mass shootings has disagreed whether attitudinal effects are present and, if they are, whether these effects are polarizing or a uniform leftward shift (Barney and Schaffner 2019; Hartman and Newman 2019; Rogowski and Tucker 2019). An absence of an attitudinal shift does not alone undermine GMAL and Yousaf's results—after all, attitudes aren't behaviors—but it provides a theoretical reason to question this result. Ultimately, however, our goal here is to try and settle the first-order question of whether gun violence has any effect on vote shares in the communities in which they happen. If there was, we could then proceed to adjudicate between mobilization and persuasion mechanisms. As we show, however, there is virtually no support for any meaningful effect on vote shares.

[12] We are grateful to each of the author teams because in each case, we were successfully able to replicate all their reported findings using their models and code.

[13] In short, HHB focus on school shootings occurring between 2006 and 2014 and, in robustness checks, between 2000 and 2018, GMAL focus "rampage-style" shootings between 1980 and 2016, and Yousaf uses the FBI's definition of a mass shooting "leading to four or more deaths at one location or crime scene" with data from 2000 to 2016.



APSR Submission Template

coding are ultimately not what drives divergent findings across studies on the topic.[14] Hence, we focus on the methodological approaches in the rest of this paper.

## 3A. EFFECT SIZES

Below in our samples, we are interested in not only the statistical significance of the effects estimated, but also their magnitude. While whether an effect is small, medium, or large or somewhere in between is always somewhat in the eye of the beholder, we use several tools to quantify the size of our observed effects. First, we benchmark our estimates to those provided by other similar geographic-based treatments in the same sample that we employ here. Second, we use equivalence testing to see what effects we are able to rule out. Equivalence testing starts from the perspective that the effects are sizable and then looks for sufficient statistical evidence that this is not the case (Hartman and Hidalgo 2018). In essence, it changes the statistical test from a difference from 0 (the standard in null-hypothesis significance testing) to test for a sizable difference—one determined by researchers. In practice, this is often done by using the confidence intervals from an estimated effect and determining what effect sizes can confidently be ruled out—using the upper and lower bounds allowed by the confidence intervals. Finally, we note that though statistical significance does not capture the full scope of the size of effects, when our effects are not statistically significant we note that this is the case.

## 4. THE TWO-WAY FIXED EFFECTS ESTIMATOR IN OUR EMPIRICAL CASE

As we detailed previously, the most common approach to estimating difference-in-differences effects when the treatment varies over time and space is the two-way fixed effects estimator (TWFE). This

[14] Another source of researcher degrees of freedom, and another difference between the three papers, is the outcome measured. GMAL only focus on presidential elections. Yousaf includes gubernatorial, Senate, and House returns, but only from Presidential election years. However, for reasons unexplained, in Yousaf's data there are only 1,715 observations in the House elections and 95.1% of the years are coded as missing. Hence, the model for House elections omits year fixed effects—making it not a true TWFE estimator. Hence, when analyzing Yousef's data, we focus on Senate and Gubernatorial elections. HHB look at a much broader set of outcomes, including presidential, congressional, state, and local elections. Ultimately, however, this difference too does not explain the difference in results. Rather, the differences stem from differences in model specifications. As shown below, models finding significant effects are biased by violations of parallel time-trends assumptions.



APSR Submission Template

approach is followed by GMAL and Yousaf, who rely exclusively on county and year fixed effects in their model specifications.[15] We replicate this approach with the data provided by HHB although this is not their primary estimator.[16] The TWFE is specified in Equation (1). For our applied example, $Y_{ct}$ represents Democratic vote share in a given county (c) and election period (t), $\phi_c$ represents a county fixed effect, $\lambda_t$ is a year fixed effect, $\epsilon_{ct}$ is the error term, and $D_{ct}$ denotes the treatment—that is, whether a county (c) in a given election period (t) is exposed to a mass shooting. $\beta$ then is the effect of interest—the TWFE of the effect of mass shootings on Democratic vote share.[17]

$$Y_{ct} = \phi_c + \lambda_t + \beta * D_{ct} + \epsilon_{ct} \tag{1}$$

With this TWFE estimator, another researcher degree of freedom exists regarding the exact nature

[15]These papers also include some time varying controls, but the bulk of identifying assumptions come from the county and year fixed effects. In some specifications, Yousaf compares successful shootings with non-successful shootings and in others includes flexible population time trends. GMAL, in some specifications, use neighboring counties as the control group, state fixed effects instead of county fixed effects, or decade fixed effects as opposed to year fixed effects.

[16]HHB include unit-specific time trends.

[17]Another researcher degree of freedom (of less consequence here) is how to estimate standard errors. If treatment assignment is completely independent across counties, you would only need to cluster at the county level (Abadie et al. 2017). However, if treatment assignment is correlated across counties—i.e., when one county in a state gets a mass shooting, it is more likely that another county gets a mass shooting—then results from the econometrics literature suggest one would want to cluster at the state level (Abadie et al. 2017). While it may seem like a better approach to simply cluster at the higher level; this is not a given. Clustering at higher levels than the treatment put researchers in the position where "to be conservative and avoid bias and to use bigger and more aggregate clusters when possible, up to and including the point at which there is concern about having too few clusters" (Cameron and Miller 2015, 333). Indeed, Abadie et al. (2017, 1) show "there is in fact harm in clustering at too aggregate a level." Given that there a very limited number of states, and thus a very real concern of having too few clusters (Cameron and Miller 2015), we cluster all of the results at the level of the treatment (the county level). Ultimately, however, once models are adjusted for time trends the clustering decision is of less consequence. This may not be true in all applications, so we advise caution and thoughtfulness in how to cluster one's standard errors.



APSR Submission Template

of the treatment.[18] One possibility is to code units exposed to treatment in a given period as being treated and all other observations—pre- and post-treatment in eventually-treated units and those who are never treated—as part of the control group. In the mass shooting example, this approach codes all counties with a mass shooting in a given electoral cycle as being treated, but county-level observations before and after that electoral cycle (along with those who never have a shooting) as not having been treated. This approach allows both in- and out-switchers in the treatment. Theoretically, this approach assumes that any effects of mass shootings would be constrained to the immediate electoral cycle only. Alternatively, another approach is to code treatment so all observations before a shooting occurs would be in the control group (along with never treated observations), but all observations in treated units post-treatment coded as treated. In our example, this approach codes all counties with mass shootings in a election cycle and election cycles that follow as treated, and all counties before—along with those who never have a mass shooting—as untreated. This means there are no out-switchers. This approach allows mass shootings to have longer term effects, changing the electoral environment both when they occur and afterward.[19]

Both approaches are used in practice when estimating difference-in-differences models. The choice between the two is a researcher degree of freedom and should be motivated by theory. Here, given a lack of strong expectations about the temporal effects of mass shootings, rather than rely on one of these assumptions, we use both. (We complement these approaches with an event-study design described below that explicitly models effects in periods before and after shootings with lags and leads.)

[18] As HHB note, the coding of treatment doesn't need to be constrained to the county in which a shooting occurs. As such, HHB examine (and fail to find) effects in surrounding counties, as a function of the distance to a shooting, as a function of the severity of a shooting, and at the national level (with daily voter registration counts as the outcome). Perhaps because they find effects at the county-period level, GMAL and Yousaf do not consider alternate treatment codings. In this paper, we focus on the shooting in a given county period given that this is the specification where GMAL and Yousaf argue an effect arises.

[19] Figures S13 and S14 provide visual illustrations of these two approaches (for a random sample of the observations) using the `panelView` package developed by Mou et al. (2022a).



APSR Submission Template

## 4a. Two-Way Fixed Effect Estimates of Shootings' Effect on Vote Shares

We start by showing that different conclusions across studies on the electoral effects of shootings are not driven by data choices. Figure 1 removes differences in methodological approaches in previous studies and shows the two-way fixed effects models (the methodological choice of GMAL and Yousaf) using the data from all of the three studies on the electoral effects of gun violence. Figure 1 also splits the results by treatment coding approaches outlined in the last section.

As shown in Figure 1, TWFE specifications consistently produce substantive positive statistically significant effects regardless of the time-frame, what shootings count as treatment—be they "rampage-style" school shootings (GMAL), school shootings (HHB), or mass shootings more broadly (Yousaf)—and regardless of how long the treatment applies. In short, TWFE estimates, regardless of the coding of shootings and treatments used, indicate a positive effect of shootings on Democratic vote share that is statistically significant and substantively meaningful.[20] Simply, when using the same model choices, shootings of different types consistently have the same modeled effect. In other words, previous differences in conclusions across studies of the electoral effects of gun violence are not due to choices about how to code shootings that might have an effect.

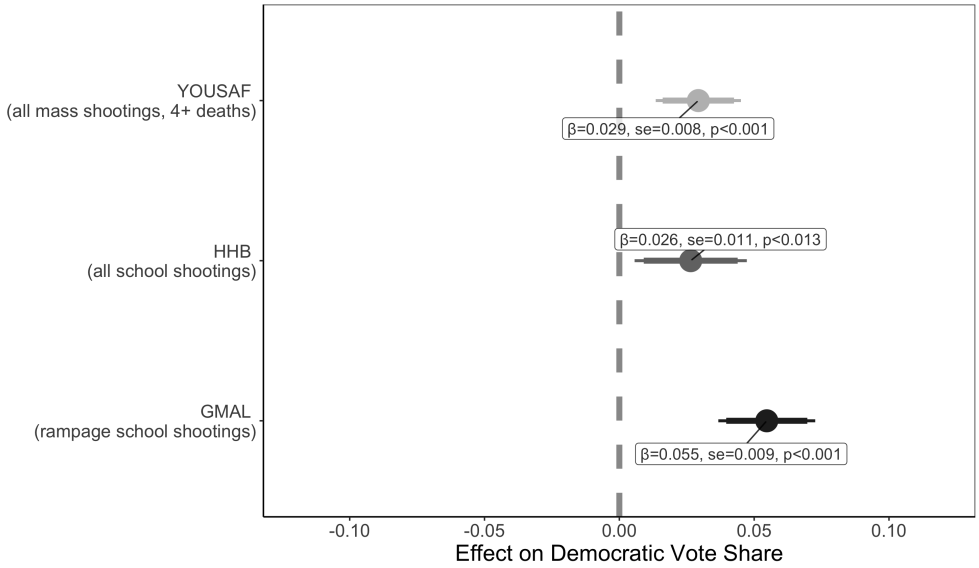
[20]Using equivalence testing (Hartman and Hidalgo 2018), in the top panel we can rule out effects smaller (larger) than 3.6 (7.3) percentage points and 6.4 (11.0) percentage points for GMAL's data (top and bottom panel, respectively), rule out effects smaller (larger) than 1.3 (4.5) percentage points and 4.8 (6.8) percentage points for Yousaf's data (top and bottom panel, respectively), and effects smaller (larger) than 0.5 (4.7) percentage points and 4.7 (10.7) percentage points for HHB's data (top and bottom panel, respectively)



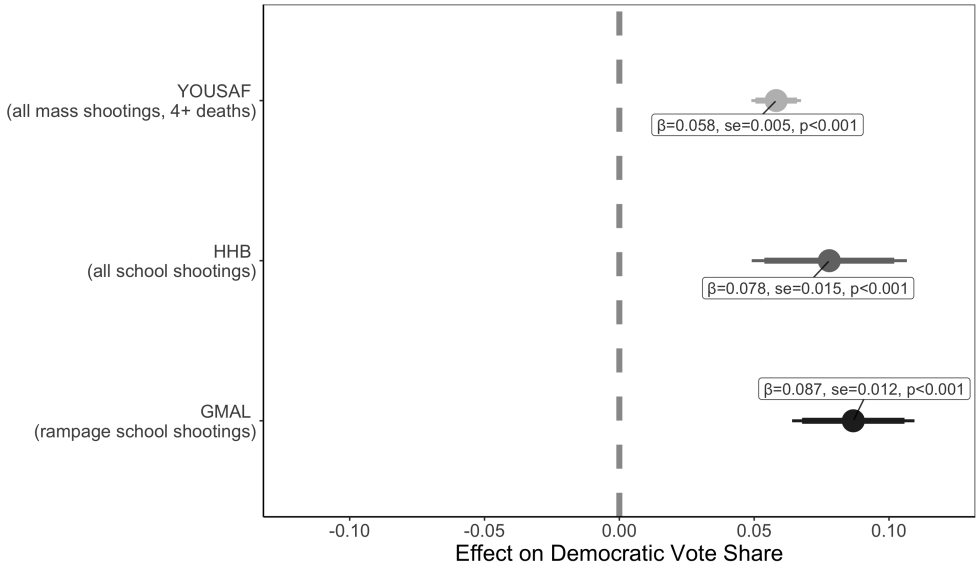
APSR Submission Template

**FIGURE 1. Differences in Previous Studies' Estimated Effect of Mass Shootings on Election Outcomes Are Not Driven by Data Choices**

**(a) Only Elections in Counties in Election Cycle When Shootings Occur are Treated**




**(b) All Elections in Counties After Shootings Occur are Treated**




Effect of mass shootings of various types are very similar (be it only "rampage-style" school shootings (as in the case of GMAL) school shootings in general (as in HHB) or mass shootings not restricted to school grounds (as in Yousaf). Estimates include county and year fixed effects (i.e. the TWFE estimator) with standard errors clustered at the county level—i.e. where treatment occurs. The top panel shows effect estimates coding only the election immediately after a shooting occurs as having been treated; the bottom panel considers all counties with a shooting treated if the election occurs after the shooting did. Model specifications for the top panel for GMAL parallel those in their Figure 4, with the exception of holding out controls (if controls are included the effect is 4.51 percentage points); GMAL don't run models equivalent to the bottom panel estimates. Model specifications parallel those in Yousaf Table 4, with a TWFE used to run a parallel specification across the papers. Model specifications for the top panel parallel those in HHB figure A11; HHB do not run the bottom panel specification. Coefficients, standard errors, and p-values are labeled for each coefficient. **Takeaway:** Differences in the statistical significance of the effects are not due to data choices of HHB, GMAL, or Yousaf. TWFE estimators suggest that mass shootings—regardless of the data/coding used—increase Democratic vote share in the county in which a shooting happens by 2.6-8.7 percentage points.



APSR Submission Template

We pause to discuss the magnitude of these estimates. An upwards of an 8.7 percentage point shift in Democratic vote share in a presidential election year (the point estimate derived from GMAL's data using the second coding of treatments) is large—as are many of the other estimates. As GMAL note, these statistically significant effects represent "a remarkable shift in an age of partisan polarization and close presidential elections" (GMAL 2021, 809). We can get a sense of how large the effects are by benchmarking them to other studies using the Leip Data on county level presidential vote shares and difference-in-differences designs. For example, Sides et al. (2022, 709) estimate that a six standard deviation shift in relative television advertising leads to a 0.5-point change in two-party vote share. Hence, if we believe these results travel, GMAL's simple TWFE estimates indicate one school shooting has an effect on Democratic vote share equivalent to a shift of approximately 66-104 standard deviations in relative advertising.[21] Using an economic comparison—the most common of retrospective voting treatments—Healy et al. (2017, 1423) show that a "1 percentage-point increase in mortgage delinquencies increases Democratic vote share by 0.33 percentage points." Thus, the effect of a 'rampage-style' shooting is roughly the equivalent to a 16.7 - 26.4 percentage point increase in mortgage delinquencies; or, in other words, moving from a world where no one is delinquent on their mortgages to a world where about 1/5 residents are at risk of losing their homes.

In short, the TWFE models suggests gun violence—regardless of how a shooting is coded—fundamentally reshapes electoral results in the local communities in which they occur. Is this sizable relationship, truly, causal and robust? Recent methodological developments provide us a guide to answer this question.

## 5. ADDRESSING ISSUES WITH TWO-WAY FIXED EFFECTS ESTIMATORS

Recent research has shown that simple TWFE models can be problematic for important reasons, which include:

1. violations of the parallel-trends assumption (e.g., Liu et al. 2021; Rambachan and Roth 2021;

[21] Similarly, the TWFE using HHB's data likewise suggests that school shootings have a positive effect for Democrats that is equivalent to a 31.2-93.6 standard deviation shift in the relative advertising advantage; Yousaf's TWFE estimate likewise suggests an equivalent effect of a 34.8-69.6 standard deviation shift in relative advertising.



APSR Submission Template

Freyaldenhoven et al. 2021),

2. mistaken inferences derived from heterogeneity in treatment effects (e.g., Goodman-Bacon 2021; Sun and Abraham 2021).

Here we discuss these issues in order and show how to apply the solutions articulated in the literature using the example of mass shootings' seeming effects on Democratic vote share.

## 6. ASSESSING & ADDRESSING PARALLEL TRENDS VIOLATIONS

One of the core assumptions to the difference-in-differences design is the so called parallel-trends assumption. In its most simple terms, the parallel-trends assumption asserts that the outcomes of interest from pre- to post-treatment would have moved in parallel among the treated and the non-treated group if not for the treatment. If the parallel-trends assumption is violated, estimated effects will be biased. Though this assumption is partially untestable because we do not observe treated units in the control group post-treatment (and vice-versa), if treated and untreated units are not moving together before treatment exposure, this could signal potential issues leading to bias (Marcus and Sant'Anna 2021; De Chaisemartin and D'Haultfoeuille 2022). There are several ways to assess the potential for differential pre-treatment trends. An appropriate first step is to visually inspect the data to see whether, prior to treatment, treatment areas are trending in directions different from the control. Then one can test for treatment effects in lagged periods, followed by an examination using event study designs. We discuss each of these in turn.

### 6a. Checking for Visual Evidence of Differential Pre-Treatment Trends

Figure 2 checks for differential pre-treatment trends using data on gun violence and elections separating counties into two bins; the first (Panel (a)), contains counties with a shooting setting aside all post-treatment observations, and the second (Panel (b)) contains all counties without a shooting.[22,23] This leaves a set of pre-treatment observations for each county that had a shooting in the sample. Figure 2 illuminates what TWFE models absorb and do not absorb. County fixed effects adjust for the differences

[22]Figure 2 use the GMAL data. Figure S10 shows analogous figures for Yousaf and HHB which also show different trends.

[23]For the few counties that had multiple shootings in this period, we are coding here the first shooting that occurred and treating all years after that as post-treatment observations.



APSR Submission Template

in Democratic vote share across counties (i.e. the vertical distance between the lightly-shaded lines). Year fixed effects account for differences that transcend counties within a given year; that is differences across years (i.e. the horizontal differences in the lightly shaded lines). What TWFE models *don't* account for, however, is the possibility that counties' Democratic vote shares change at different rates over time, a significant problem in the context of school shootings.

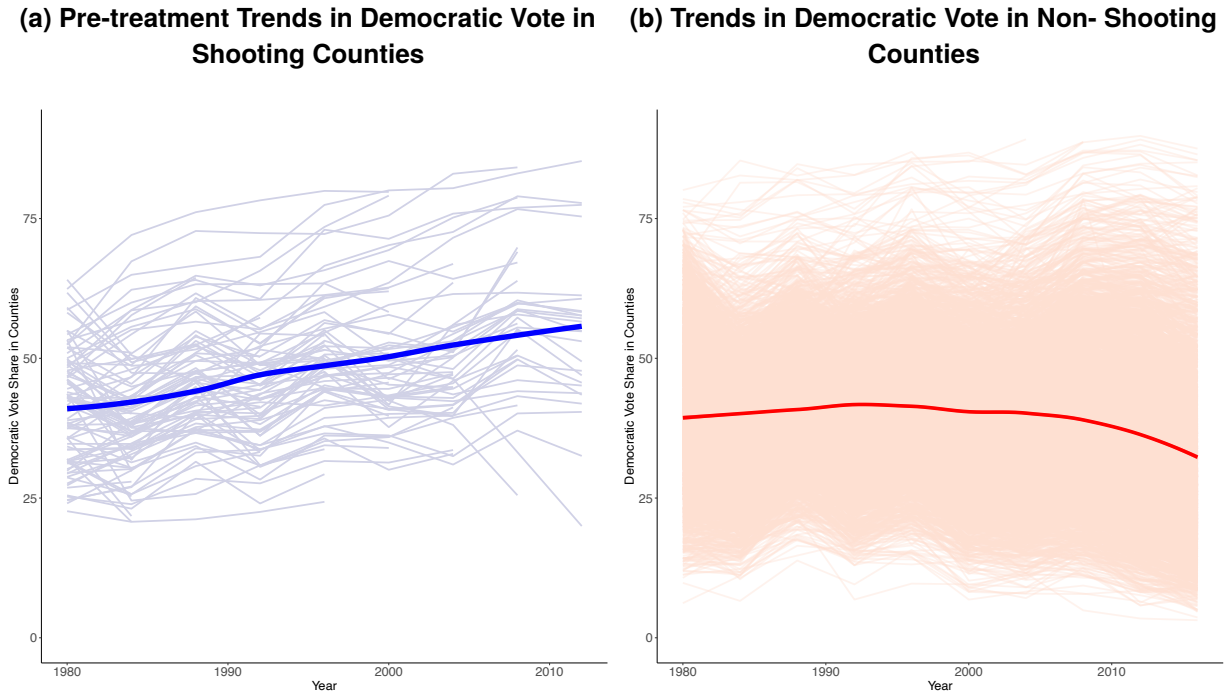
As seen in Figure 2, shootings happen in areas that are—proceeding and unrelated to shootings themselves—more likely to be trending Democratic relative to the rest of the country over the last 40 years. This is likely because mass shootings disproportionately occur in growing populations and have been increasing over time (U.S. Government Accountability Office 2020; Musu-Gillette et al. 2018) combined with a realignment in American politics in which these same more populated areas have become consistently more Democratic over time (DeSilver 2016). This provides a cautionary tale for other contexts; researchers should take great care generally in instances where demographic/political change that predates treatment aligns with short-term treatment exposure.

In our case, this (coincidental) pre-treatment trend separation between counties with and without a shooting becomes especially prevalent after 2004. This is particularly problematic because GMAL explicitly note the effects of shootings on Democratic vote share between 1980 and 2000 are essentially null (or negative) but that starting in 2004 the positive effects on Democratic vote share begin to increase (Garcia-Montoya et al. 2022, Figure 7). Figure 2 indicates this is the exact time when the parallel-trends assumption becomes particularly tenuous as areas with and without shootings move more clearly in opposite directions. Two-way fixed effects models do not absorb these trends and, as such, are likely to be biased in these circumstances.

## 6b. Checking for Pre-Treatment Effects with the Model Specifications Used

While Figure 2 shows visual evidence of a likely difference in pre-treatment trends, it is not dispositive. Researchers could tinker with the axes to minimize or exacerbate the appearance of differential trends. Hence, the next check to assess TWFE design validity—and one that should be standard practice—is to check whether this model specification suggests any impact *prior* to treatment (Grimmer et al. 2018). In our case, this placebo test is informative as shootings—something that people cannot precisely

**FIGURE 2. Trends in Presidential Vote in Counties With Mass Shootings Prior to Shootings, Compared to Trends in Counties Without Shootings**




Pre-treatment trends of Democratic vote share in counties where a shooting occurred benchmarked to the trends in Democratic vote share found in counties where a shooting did not occur. In the panel on the left, the small blue lines mark the patterns for all counties with a shooting and the bolded **blue** line captures the average trend across these counties. The panel on the right shows the same pattern for counties without a shooting. The small red lines mark the patterns for all counties without a shooting and the bolded **red** line shows a loess model for counties without a shooting. **Takeaway:** Counties that have shootings trended more Democratic *even before* the shootings occurred, whereas counties without a shooting trended slightly more Republican. The result is that models will be biased that don't account for differential trends across counties (i.e. TWFE models) will be biased.

anticipate—should not affect Democratic vote shares prior to a shooting. If there are effects, we should be skeptical the TWFE estimate is, indeed, causal (Hansen and Bowers 2008; Angrist and Pischke 2008).

Specifically, this step examines effects on lagged measures of the outcome variable by running the specification listed in Equation (2) below. Equation (2) is the same as Equation (1), except for the outcome variable. Here, instead of estimating the effect of shootings (i.e. $D_{ct}$) in the election following the shooting (i.e. $Y_{ct}$), we substitute a lagged version of the outcome variable (i.e. $Y_{ct-k}$). Here $k$ corresponds to the number of lagged periods one wishes to include. We include seven lagged periods in our models as the GMAL panel is sufficiently long to do so. However, power considerations may influence the number of lags used. We recommend scholars, as a first step, look for effects in the one period lag, then look to see how far back they can estimate precise-enough specifications for their



APSR Submission Template

applications and then estimate those specifications. As Hartman and Hidalgo (2018) note, tests for imbalance are context specific and scholars should consider how precise of an imbalance is meaningful using equivalence testing.

$$Y_{ct-k} = \phi_c + \lambda_t + \beta * D_{ct} + \epsilon_{ct} \tag{2}$$

In this case, our imbalances are statistically significant and substantively meaningful. Panel (a) and (b) (the top section) of Figure 3 show the TWFE model for the various types of shootings on lagged measures of Democratic vote share.[24] (We return to a discussion of Panels (c)-(f) below.) Regardless of treatment coding, there is substantial imbalance in lagged outcomes.[25] We start, on the left of each panel, with the presidential election 4 years prior to the shooting and work back to up to 7 presidential elections (i.e. 28 years) prior to when the shooting occurred.[26] These vary somewhat by specification, but tend to range around a 2-7 percentage point effect, with most of these being highly significant. This analysis indicates that mass shootings have a significant and substantive positive effect on Democratic vote shares up to and including 20 years *prior* to when a shooting occurred and these effects are all substantively meaningful and statistically significant. Simply, the TWFE does *not* recover balance prior to shootings, regardless of the data used.

The effects shown in panels (a) and (b) of Figure 3 should not exist if TWFE estimators were uncovering causal effects as there is little reason—theoretically or empirically documented—to suspect that school shootings should have anticipatory effects many years prior, given these events are relatively unexpected in the communities where they occur. In other applications, this may not be true; for example, in policy evaluation studies there may be reasons for anticipatory effects if a policy, say, is announced before it is implemented. However, anticipatory effects are infeasible in the mass shootings context and differences are further evidence that areas that had shootings were naturally *trending* more Democrat before a shooting occurred (indicators of a highly likely violation of the parallel-trends assumption).

[24] See Tables S13 and S14 for the Table version of the results from (a) and (b).

[25] HHB control for pre-treatment trends, but using their data with TWFE estimates also produces a 2.2 percentage point increase in Democratic vote share 4 years before a school shooting ($\beta$=0.022, $p$<0.073).

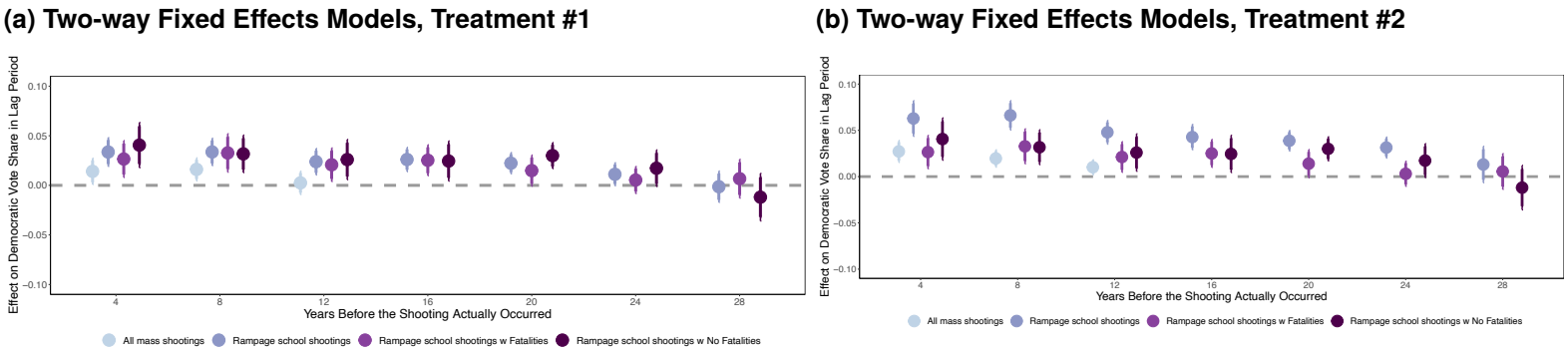
[26] Yousaf's time-frame is only long enough to look three Presidential election cycles back in time.



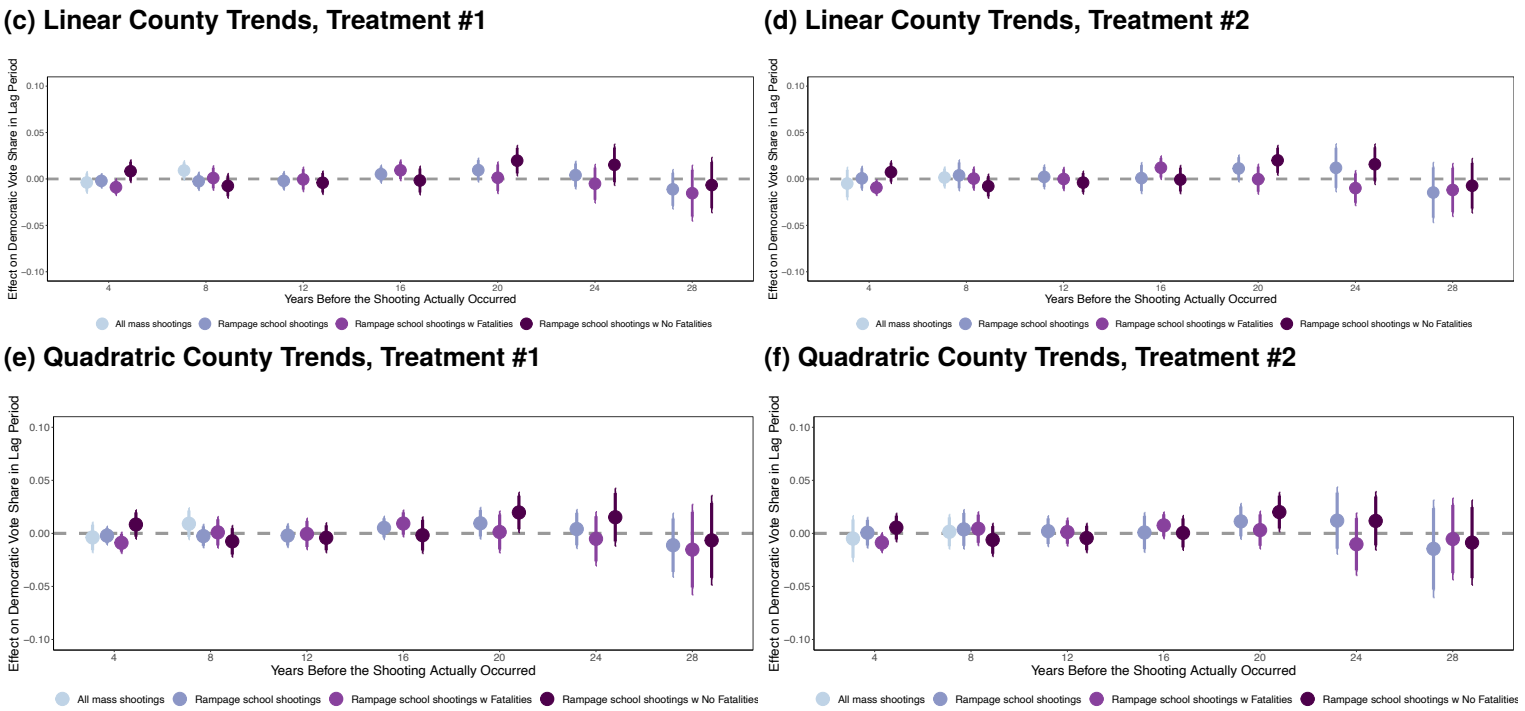
APSR Submission Template

## FIGURE 3. The Effect of Shootings on Election Outcomes Many Years Before

### Without Time Trends




### With Time Trends




Effect of mass shootings on Democratic vote share in the years prior to when a shooting occurred. Treatment #1 is coded such that only elections with shooting coded as treated; Treatment #2 is coded such that all elections after a shooting occurs in a county are coded as treated. Years prior span from 4 to 28 years. Estimates for Yousaf can only extend part of that range given the shorter time-frame used in this dataset. All estimates include county and year fixed effects, the second row adds county-specific linear time trends, the third row adds quadratic county-specific time trends. Columns show two ways of coing treatment: . All models' standard errors are clustered at the county level—i.e. where treatment occurs. Coefficients, standard errors, and p-values are labeled for each coefficient. **Takeaway:** TWFE estimators show signs of shootings having an effect up to and including 20 years prior to when a shooting occurred. This is not possible, indicating potentially fatal identification issues with this model approach. In contrast, specifications with linear and quadratic time trends show balance prior to when the shooting occurred.

APSR Submission Template

## 6c. Checking for Pre-Treatment Trends with Event-Study Designs

There is the possibility that these pretreatment effects could show up where there is not bias if treatment in one period (i.e. $D_{ct}$) was highly correlated with treatment in prior periods (i.e. $D_{c,t-k}$). In that limited case, the coefficient on $D_{ct}$ may show an effect if there is an effect of $D_{c,t-k}$ on $Y_{c,t-k}$. If this were the case, there could be pre-treatment effects even if there is absolutely nothing wrong with the research design. For this reason, we recommend not stopping at the lagged outcomes test and also model the effects of lagged and leaded treatment, using an event-study design.

In our application, another way to see the pre-treatment imbalance is to create lag and lead measures of the treatment variable and set up the model as an event-study design tracing effects before and after treatment (Binder 1998; Armitage 1995). An event-study is a type of difference-in-differences model becoming increasingly common given its less restrictive and more transparent modeling assumptions, but its usage is relatively rare in political science.[27] An event-study (usually) preserves the two-way fixed effects, but also includes a series of lagged and lead treatment variables. The event-study specification is shown in Equation (3) below, where we list treatment in a given county ($c$) and year ($t$), lagged or leaded by the corresponding periods since treatment ($k$). For simplicity, we show the event-study model for one pre-treatment period ($k-2$), the period when treatment occurs ($k$), and the period after treatment occurs ($k+1$). As in most event-study designs, the baseline is the period before treatment occurs (i.e. $k-1$) (Binder 1998; Armitage 1995).

$$Y_{ct} = \phi_c + \lambda_t + \beta_{-2} * D_{ct,k-2} + \beta_0 * D_{ct,k} + \beta_1 * D_{ct,k+1} + \epsilon_{ct} \tag{3}$$

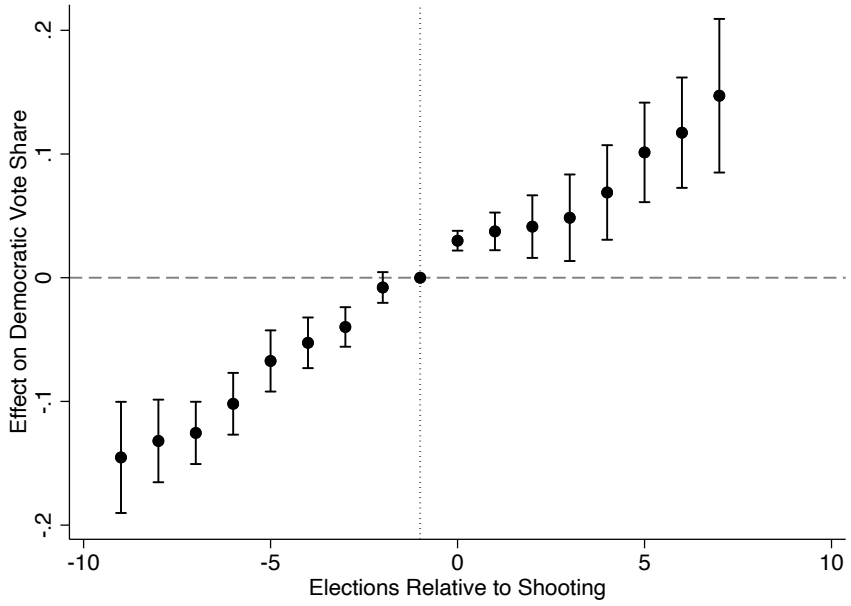
Figure 4 visualizes the event-study with nine pre-election treatments and eight post-treatment periods included.[28] (Table S18 in the Online Appendix provides the estimates that produce Figure 4.) As shown on the right of Figure 4 (i.e. right of the grey dotted vertical line), there is an immediate, significant, and substantive jump in Democratic vote share in the election year when a shooting occurs. The TWFE event-study model also implies that the effect of a single rampage-style school shooting grows after the

[27] A Google Scholar search of articles in the *American Political Science Review* yields only 11 total articles that mention an event-study design.

[28] Here we focus on the GMAL data because this approach requires sufficient observations before and after treatment and GMAL's data is the longest of the three studies.

event occurs, still having an increasingly larger effect on Democratic vote share (10-13 points). These effects are statistically significant, substantively meaningful, and allow us to rule out smaller effects using equivalence testing. While not completely impossible—as opposed to the pre-treatment effects documented earlier—it is long-lasting, and the increasing effect remains theoretically unexplained.

**FIGURE 4. Event-study Estimates Show that TWFE Fails to Account for Pre-Treatment Trends**




Event-study estimates with county and year fixed effects (GMAL data). Baseline election year is shown with a grey dotted line; as is common, the baseline is one-year prior to when a shooting occurs. Those to the left are pre-treatment years and those to the right are post-treatment years. Following prior work, we bin our extreme points (Schmidheiny and Siegloch 2019; Baker et al. 2022); in practice, this means that elections 8 and 9 are binned with election 7 given small sample sizes in these outer bins. The grey dashed line is a reference line for a zero effect. Points are point estimates for lags/leads and bars are 95% confidence intervals. **Takeaway:** If one only looks at the differences in Democratic vote share between counties that had shootings and those that did not in the years after the shooting, one would conclude that shootings fundamentally change elections. However, the left of the panel shows that counties that have shootings trended more Democratic *even before* the shootings occurred. The increase that occurs after a shooting is entirely consistent with a general trend towards more Democratic election outcomes. The result is that models that don't account for differential trends across counties—even if they are done in an event-study framework—will be biased.

However, Figure 4 also illuminates the "effect" documented by the TWFE models is unlikely to be causal as indicative by looking to the left of the baseline period (i.e. left of the grey dotted vertical line). If the TWFE models were causal, these coefficients should not be significantly/substantively distinct from zero. This is *not* what we observe. Relative to one election prior to when shootings occur, prior year elections see lower support for Democratic candidates and this underperformance increases as we move back in time. In other words, vote share trends more Democratic prior to shootings in



APSR Submission Template

counties where shootings occur (relative to counties where shootings do not occur). What happens after a shooting is just a continuation of that trend—as the points themselves are almost a perfect linear function with elections relative to shootings—and further evidence that the TWFE estimators are biased in the case of mass shootings and electoral outcomes. Given the value of this approach in more formally testing for pre-treatment imbalances, we recommend that parameterizing models as an event-study also become standard in difference-in-differences applications.

## 6d. Controling for Any Differential Unit-Specific Time Trends

Facing potential violations of the parallel-trends assumption, one potential (and often straightforward) solution is to adjust for factors—observed or unobserved—leading to pre-treatment imbalances. In this case, visual inspection (see Figure 2) reveals that treated and untreated units trend in different directions. A solution is to include unit-specific time trends—in this specific case, count and year time trends controlling for differential trends in Democratic vote share (Wing et al. 2018; Angrist and Pischke 2008, 2010). The identifying assumption becomes the deviation from county-year trends captured by the interaction of time with each unit. Identification comes from sharp deviations from otherwise smooth unit-specific trends.

This specification corresponds Equation (4) which adds a fixed effect for each county ($c$) interacted with time ($t$) in the estimation.

$$Y_{ct-k} = \phi_c + \lambda_t + \gamma_{c*t} + \beta * D_{ct} + \epsilon_{ct} \tag{4}$$

While Equation (4) includes linear county-specific time trends—$\gamma_{c*t}$, the functional form of the trends included is a potentially influential researcher degree of freedom.[29] Modeling county-specific trends incorrectly could lead to mistaken inferences. As a result, we run a host of model specifications—all of which take slightly different tacts to adjusting for differential pre-trends. In the next section, we also show various other approaches to adjusting for differential pre-trends, including the methods recently developed by Liu et al. (2021), Freyaldenhoven et al. (2021), and Rambachan and Roth (2021),

[29]In another approach, we change the dependent variable to the change in Democratic vote share from the election before shootings occurred to the election in which counties are actually treated. This approach helps skirt the so-called Nickell bias that arise when considering models with lags and fixed effects (Beck et al. 2014).



APSR Submission Template

described below. We recommend scholars consider running the check of robustness across several parameterizations of the unit-specific trends, acknowledging higher-order unit specific trends could face a bias-variance tradeoff, especially in smaller datasets.

Figure 3 (panel (c)-(f)) show the pre-treatment effect models with linear and quadratic (i.e. adding $\gamma_{c*t^2}$ to Equation (4)) trends. We also include cubic and quartic county-specific time trends in the Online Appendix (see Figures S9 and S10). In contrast to the TWFE (panels (a)-(b) in Figure 3), all models with county-specific trends (panels (c)-(f) in Figure 3) are balanced pre-treatment. Importantly, these null effects (and especially those in more proximate periods) are not driven by an inflation of the standard errors and allow us to precisely rule out even very modest pre-treatment differences using equivalence testing. For example, 4 years prior to a rampage style shooting, in the linear trends model (treatment #1) we can rule out pre-treatment effects smaller than -0.99 percentage points and effects larger than 0.57 percentage points. These effects are much smaller and distinct from the pre-treatment effects we see in the TWFE.

Figure 5 shows estimates for models including linear and quadratic county-specific time trends on the post-treatment outcomes for the first treatment coding (i.e. only counties with shootings in that year coded as treated).[30] Figure S14 in the Online Appendix shows the results for treatment #2 (i.e. counties with shootings do not revert to the control afterwards). As Figure 5 shows, once we make this necessary adjustment, the effects of mass shootings—be they at school or 'rampage-style' only—attenuate heavily. All of the effect estimates are *much* smaller than the original estimates and virtually all are no longer statistically significant at traditional levels.

Specifically, panels (a) and (b) of Figure 5—the effect estimates for rampage style school shootings (i.e. GMAL's treatment)—indicate a 0.7 percentage point increase in Democratic vote share, an effect that is no longer statistically significant but still fairly precise. Substantively speaking, this effect is 7.9 times—i.e. 790%—smaller than the original TWFE.[31] Using equivalence testing, we can rule out effects as large as the TWFE with a high degree of confidence. For example, in the rampage school shootings treatment (i.e. GMAL's treatment) with linear trends, we can confidently rule out effects larger than 1.5 percentage points. The same holds for the Yousaf data, which are no longer in the 3-6

[30] See Tables S19-S22 for the model estimates that produce these figures.

[31] Using the second treatment coding, the effect is 7.3 times smaller than the original TWFE.



APSR Submission Template

percentage point range, but now hover almost exactly at zero.[32] None of these effects are statistically significant once trends are added ($p > 0.77$ in all specifications). And using equivalence testing based on the estimate from the linear trends model we can rule out effects larger than 1.18 percentage points. Finally, all of the effects using HHB's data are very close to zero, not significant, and powered sufficient to allow us to rule out very modest effects using equivalence testing. Simply, regardless of the data used, there is no clear evidence of the *large* effects documented in previous work finding an effect, and no consistent evidence for effects that are statistically distinguishable from zero.[33] While statistically significant effects infrequently show up in Figure 5, they are not robust. Notably, if we add higher-order polynomials—-as we do in Figure S9 in the Online Appendix—no effects are significant. This—along with further checks below—bolsters the argument that we cannot support the conclusion that shootings have significant and meaningful effects on Democratic vote shares.[34]

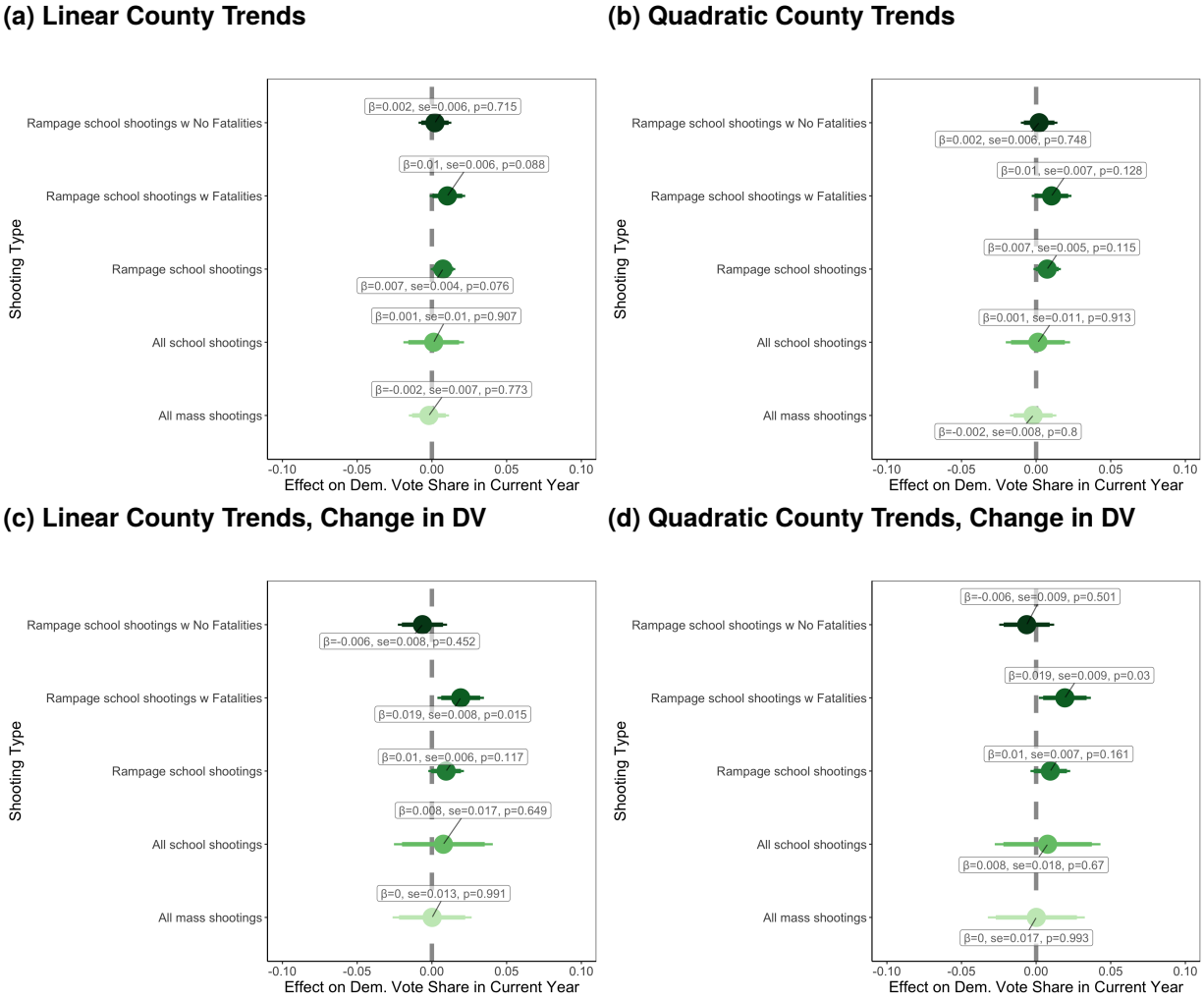
Adding leads and lags of the treatment values and using an event-study design that accounts for the

[32] As shown in Figure S14, Yousaf's effects are not significant in 3/4 of the models run with the second treatment coding.

[33] One concern might be that adding time trends adding time trends artificially inflates our standard errors to levels that are unpalatable. That said, confidence intervals remain relatively small in models with linear time trends. They are slightly less precise with quadratic time trends, but the confidence intervals are, to our eye, still quite tight with this specification and the cubic/quartic specifications shown in the Appendix.

[34] Moreover, the null effects (once accounting for time trends) continue in robustness checks run by GMAL and Yousaf. In Yousaf's original comparison of shootings versus failed shootings, estimates range from 1.2 to 2.5 percentage point gain for Democrats ($p < 0.05$ in all specifications). As HHB note, there are reasons to count neighbor counties as treated counties or consider them partially treated as a function of distance to the shooting. However, once linear or quadratic time trends are added, effect estimates are only 0.04 to 1.4 percentage points, with none close to statistical significance. With only year and county fixed effect's GMAL's original estimates using neighboring counties as the control indicate a 4.3 percentage point gain for Democrats from a 'rampage-style' shooting. Once linear ($\beta$ = 0.5 percentage points; $p<0.61$) or quadratic time trends ($\beta$ = 0.5 percentage points; $p<0.64$) are added, this effect attenuates and becomes not distinct from zero. GMAL also run model specifications with state and decade fixed effects. However, it is not clear why these models should be treated as good as models with county fixed effects given they do not absorb county-specific factors. However, effects also attenuate dramatically if we add trends to these models.

**FIGURE 5. Effects of Mass Shootings on Presidential Elections After Absorbing County-Specific Trends**




Effect of mass shootings of various types once we account for differential trends in Democratic vote share across counties in the United States. Within each panel, the first 3 estimates are using the GMAL coding of mass shootings and their data, the next comes from HHB, and the last comes from Yousaf. The upper left panel shows specifications with linear county trends, the upper right panel shows specifications with quadratic county trends, the bottom left panel shows specifications with linear county trends and using a change in Democratic vote share over the prior 4-year-previous election, the bottom right panel shows specifications with quadratic county trends and using a change in Democratic vote share over the prior 4-year-previous election. Coefficients, standard errors, and p-values are labeled for each coefficient. Cubic and quartic specifications, see Figure S9. For effects where we code all post-shooting counties as being treated—not just counties and years with shootings—see Figure S14. **Takeaway:** Once we account for differential trends across counties, the effects of mass shootings—be they located on school grounds or not, or be they rampage style or not—are all smaller and precisely-estimated.

differential pre-treatment trends bolsters these conclusions as effects are even smaller and more precise. Figure 6 uses the suggestions developed by Freyaldenhoven et al. (2021) to display event study designs and account for pre-trends in event-study designs.[35] (Tables S23-S26 show the coefficients that produce

[35]This approach uses the treatment variable that codes treatment only in the time period.



APSR Submission Template

these figures.) Figure 6 uses the same y-axis as Figure 4 for ease in comparing across the two. Once trends are added, the pre-treatment imbalances identified by the TWFE model are heavily attenuated. So too, however, do any meaningful post-treatment differences. Rather than a 5-13 percentage point effect, the immediate effect of 'rampage-style' school shootings is a 0.8 percentage point bump for Democrats. This effect is just statistically significant at the unadjusted levels in the linear trends models (p=0.048). However, this estimate still allows us to confidently rule out effects of the variety estimated by GMAl; we can rule out effects larger than 1.67 percentage points using equivalence testing. This effect is not significant at the 5% level, but is at the 10% level, in the quadratic trends model (p<0.066); in this model specification we can rule out effects larger than 1.71 percentage points.[36] These effects appear to be the upper bound produced from this method. If we use a slightly different approach to estimating the event-study with trends—that developed by Clarke and Tapia-Schythe (2021) and the corresponding `eventdd` command in `STATA`—we get estimates that negative (albeit statistically indistinguishable from zero). With this slightly different approach, the effect in the first period following the shooting is -0.13 percentage points, with a p-value of 0.898 and a 95% confidence interval that spans from -2.05 to 1.80 percentage points.[37] Moreover, the results are not robust to alternate shooting codings; the effect for the HHB data/coding is just 0.36 percentage points (p = 0.745; 95% CI: [-1.8, 2.5]). Moreover, if we test the robustness of these effects to other pre-treatment periods—as the approach designed by Freyaldenhoven et al. (2021) allows—the effects are even smaller and even less suggestive of an effect (see Figures S2-S5 in the Appendix). Benchmarked to the two-period lag trend, the estimate in first election after a shooting for the linear county trends model is a mere 0.47 percentage points (p=0.324); the 95% confidence intervals for this estimate span from very modest negative effects to very modest positive effects (95% CI: [-0.5, 1.4]). Benchmarked to the two-period lag trend, the estimate in first election after a shooting for the quadratic county trends model is a mere 0.5 percentage points (p=0.359); the 95% confidence intervals span from very modest negative effects to very modest positive

[36] In the GMAL data, most of the evidence for an effect shows up in the rampage-style shootings with killings (*beta*=1.0 percentage points; p=0.064; 95% CI: [-0.05, 2.1]) as opposed to rampage-style shootings without killings (*beta*=0.45 percentage points; p=0.550; 95% CI: [-1.0, 1.9]).

[37] Estimates from the linear trends model; those with a quadratic county-specific trend are: -0.09 percentage points, with a p-value of 0.927 and a 95% confidence interval that spans from -2.1 to 1.90 percentage points.

effects (95% CI: [-0.5, 1.5]). Moreover, none of the large longer-term effects—i.e. those beyond the first year—remain in these event study models.

In short, in an event-study that adjust for unit-specific trends, there is little evidence for an effect of the size suggested by the TWFE in the election following a shooting. In fact, there is little evidence for any significant effect. The intermittent effects that do cross the $p < 0.05$ threshold are not robust to reasonable model variations, such as the baseline comparison points one uses. And the estimates are fairly precise; allowing us to rule out very modest effect sizes using equivalence testing. In most specifications, our 95% confidence intervals include very modest negative and very modest positive effect estimates.

## 6e. Implementing Additional Checks that Address Potential Violations of Parallel-Trends Assumptions

Including unit-specific time trends, as we have done above, is not the only solution to violations of the parallel-trends assumption; nor is it likely to be the solution in all applications as scholars may desire a more flexible solution. Recent advances in the difference-in-differences literature have suggested alternative solutions to potential violations of parallel-trends assumptions or the presence of unobserved time-varying confounders. We recommend that scholars implement, at minimum, checks suggested by Liu et al. (2021) and Rambachan and Roth (2021), as outlined below.

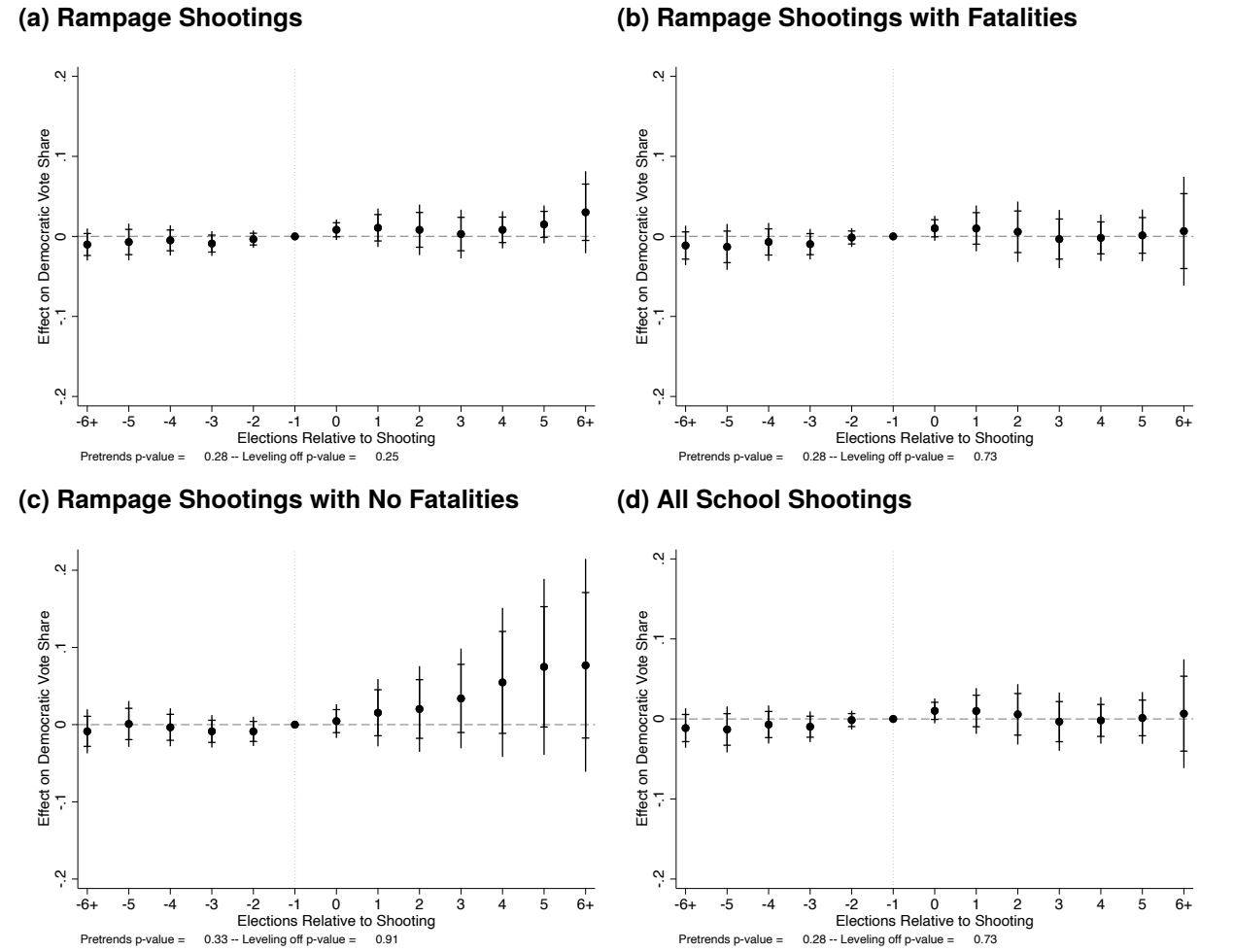
Liu et al. (2021) develop procedures—including what they call the fixed effects counterfactual estimator, the interactive fixed effects counterfactual estimator, and the matrix completion estimator—to "estimate the average treatment effect on the treated by directly imputing counterfactual outcomes for treated observations" (1).[38] Using simulations, Liu et al. (2021) show that the interactive fixed effects counterfactual estimator provides more reliable causal estimates than conventional TWFE models when unobserved time-varying confounders exist. The interactive fixed effects counterfactual estimator can be applied with the package `panelView`, which is available in both `Stata` and `R` and allows for estimation of their various estimators and dynamic treatment effects plots (Mou et al. 2022b,b). As stated in their `Stata` package, "these estimators first impute counterfactuals for each

[38]This approach uses the treatment variable that codes treatment in all treated units post-treatment as having been exposed.



APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

**FIGURE 6. Event-study Estimates of Shootings After Absorbing County-Specific Trends**




Event-study estimates from the HHB and GMAL data with county and year fixed effects and county-specific quadratic time trends. These use the method developed by Freyaldenhoven et al. (2021) to account for pre-trends in event-study designs. Analysis executed using the `xtevent` and `xteventplot` commands in `STATA` (Freyaldenhoven et al. 2022). These commands, as a default, plot both the standard confidence intervals and those developed by Montiel Olea and Plagborg-Møller (2019), which were developed for contexts with dynamic effects. As is common, the baseline is one-year prior to when a shooting occurs and is shown with a gray dotted line; other year baseline comparisons are shown in the Supplementary Appendix (see Figures S1-S4). Those to the left are pre-treatment years and those to the right are post-treatment years. The grey dashed line is a reference line for a zero effect. Points are point estimates for lags/leads and bars are 95% confidence intervals. Figure uses the same y-axis as Figure 4 for ease in comparing across the two. The bottom row shows effects with and without fatalities; these models use quadratic county trends **Takeaway:** Once time trends are taken into account in event-study models, the effect of shootings attenuates considerably.

treated observation in a TSCS dataset by fitting an outcome model using the untreated observations. They then estimate the individualistic treatment effect for each treated observation by subtracting the predicted counterfactual outcome from its observed outcome. Finally, the average treatment effect on the treated (ATT) or period-specific ATTs are calculated." Their work builds on research exploring factor-augmented models for applications surrounding causal identification (Bai 2009; Bai and Ng



APSR Submission Template

2021; Gobillon and Magnac 2016; Xu 2022).

Figure 7 applies the interactive fixed effects counterfactual estimator to our example using GMAL data. (See also Tables S27-S30.) Panel (a) shows the TWFE and panels (b)-(d) show interactive fixed effects counterfactual estimators. In the TWFE model, there are pre-treatment imbalances and a general overall upward trend—as shown in Figure 5 previously. Again, this suggests that the TWFE may be picking up on a general trend towards more Democratic votes in pre-treatment periods. However, once pre-treatment differences are adjusted for in models specifically developed by Liu et al. (2021) to address pretreatment imbalances, the overall evidence again does not support the argument that shootings substantially or significantly affect vote shares in the years following shootings. Again, this is not for a lack of statistical power. All effects are comparatively modest and the confidence intervals are precise enough to rule out very modest effects using equivalence testing. Moreover, any (smaller) effects that appear intermittently are not robust to reasonable model variations in researcher degrees of freedom.

Rambachan and Roth (2021) propose another solution to situations where parallel-trends assumptions are unlikely to hold.[39] Rather than forcing researchers to decide a rigid and exact functional form, Rambachan and Roth (2021) propose a sensitivity analysis approach to potential violations of the parallel-trends assumption. Rambachan and Roth (2021)'s sensitivity analysis avoids researchers having to arbitrarily choose a parametric model for the violations of pre-trends. This sensitivity approach is particularly useful as researchers may often struggle to know the functional form of the underlying system they are studying.

This sensitivity analysis can be formalized in three ways. First, researchers can choose to see how robust their effect is to (unobserved) post-treatment departures of parallel trends by bench-marking to the (observed) maximum pre-treatment violation of parallel trends. They call this the $RM$ approach, which stands for for "relative magnitudes." In this approach, researchers choose different values of $\overline{M}$, which measures how much of the maximum pre-treatment violation of parallel trends would lead the effects to include null effects in the confidence set. Rambachan and Roth argue that this approach is reasonable, for example, if "the researcher suspects that possible violations of parallel trends are driven

[39]This approach uses the treatment variable that codes treatment only in the time period.



APSR Submission Template

**FIGURE 7. Liu et al. (2021) Interactive Fixed Effects Counterfactual Estimator**

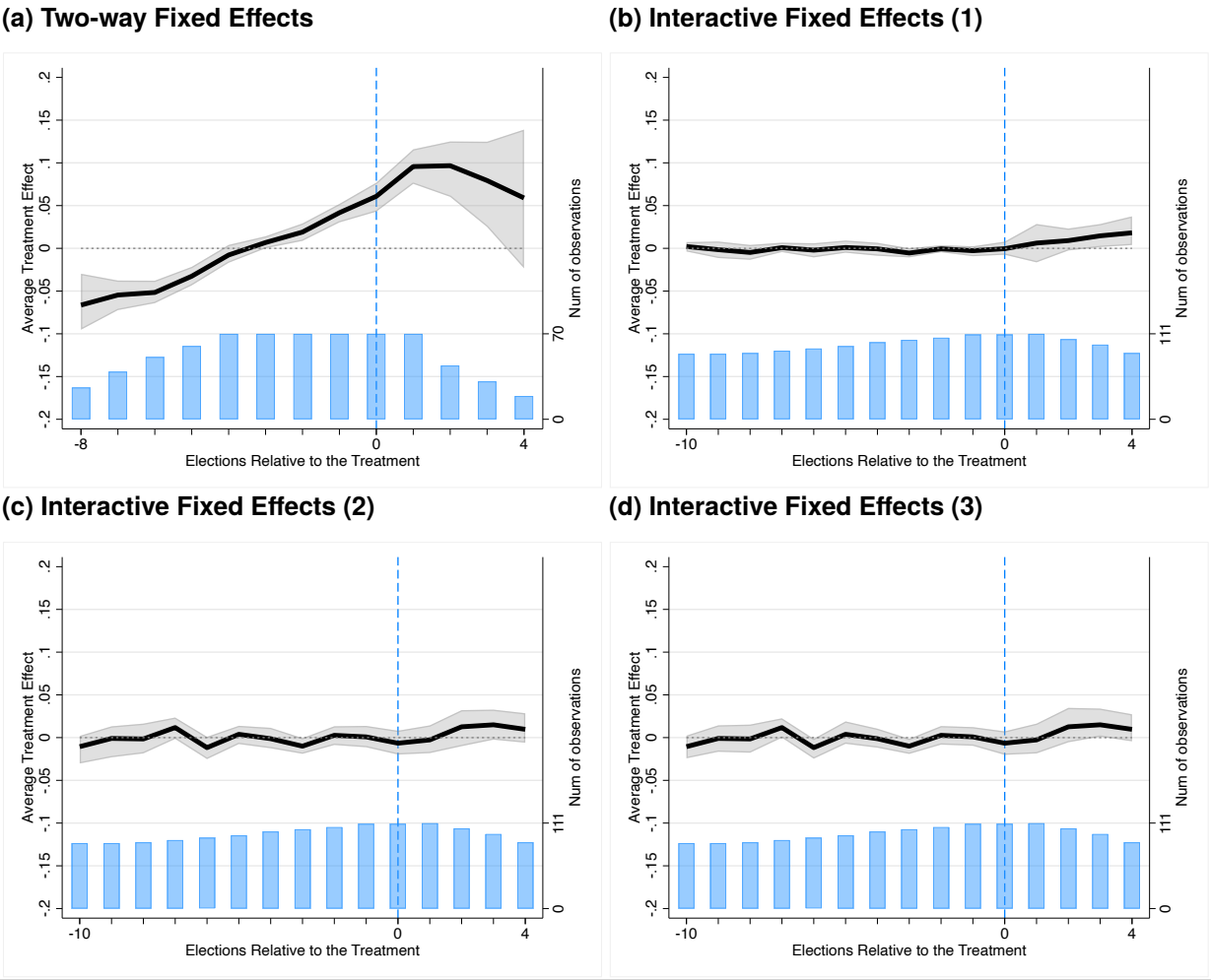



Figure shows the interactive fixed effects counterfactual estimator developed by Liu et al. (2021) using GMAL's data. The top left panel shows the TWFE estimated by Liu et al.'s (2021) `FECT` package; it is analogous to Figure 5, but their procedure estimates slight differences in the number of pre- and post-treatment periods. In the top right panel, the number of factors ($r$) is set to 3—that chosen by cross validation and the degree of the polynomial is set to 4. In the bottom row, $r$ is set to 1 in both panels and degree 2 and 4, from left to right. For other variations, see the Online Appendix. **Takeaway:** The upward trend in the TWFE model is indicative of violation of the parallel trends assumption; evidence that treated and untreated units were on considerably different paths pre-treatment; TWFE don't control for this. In the interactive fixed effects models, there is no evidence of the substantial effects shown in more simplistic model specifications that do not account for potential violations of the parallel-trends assumption.

by confounding...shocks that are of a similar magnitude to confounding...shocks in the pre-period" (Rambachan and Roth 2021, 12). Second, researchers can choose to see how robust their effect is to varying departures from differential trends evolving smoothly over time. This may be especially useful when "researchers may be worried about confounding from secular trends (e.g. long-run changes in labor supply) that they suspect evolve smoothly over time" (Rambachan and Roth 2021, 13). This sensitivity test is "done by bounding the extent to which the slope may change across consecutive

periods" (Rambachan and Roth 2021, 12). Under this approach, "the parameter $M$ governs the amount by which the slope...can change between consecutive periods, and thus bounds the discrete analog of the second derivative" (Rambachan and Roth 2021, 13). They call this the $SD$ approach, which stands for for "second derivative" and/or "second differences." Finally, researchers can combine these two approaches in the $SDRM$ condition. This approach "assume[s] that the possible non-linearities in the post-treatment difference in trends are bounded by the observed non-linearities in the pre-treatment difference in trends" (Rambachan and Roth 2021, 13). Under this approach, $\overline{M}$ is the parameter the research varies, which allows the researcher to set "bounds [for] the maximum deviation from a linear trend in the post-treatment period by $\overline{M} \geq 0$ times the equivalent maximum in the pre-treatment period." Rambachan and Roth note that $SDRM$ is similar to $SD$, "except it allows the magnitude of the possible non-linearity to explicitly depend on the observed pre-trends" (Rambachan and Roth 2021, 13).

With Rambachan and Roth's (2021) general approach, the conclusions of difference-in-differences specifications do not depend on arbitrary choices of model specification. Rambachan and Roth's approach relaxes the strong parametric assumptions behind county-level linear or quadratic trends by bounding how much the trend can deviate from linearity and/or bounding the maximum violation of parallel-trends by the maximum pre-treatment violation. In essence, this approach "show[s] what causal conclusions can be drawn under various restrictions on the possible violations of the parallel-trends assumption" (Rambachan and Roth 2021, 1). This approach is implementable through the `HonestDID` package in `R` Rambachan and Roth (2021). They articulate their approach and provide guidance for its execution here. There is not currently a corresponding `HonestDID` package in *STATA*.

Rambachan and Roth (2021, 28) note "it is natural to report both the sensitivity of the researcher's causal conclusion to the choice of this parameter and the 'breakdown' parameter value at which particular hypotheses of interest can no longer be rejected." We do this in Figure 8 below using GMAL's data. Our analyses replicate Rambachan and Roth's three different sensitivity approaches outlined above. The top left panel of Figure 8 shows a sensitivity analysis for the $SD$ approach by plotting robust confidence sets for the treatment effect in the mass shooting case for different values of the parameter $M$. The confidence sets show that the effect of mass shootings on Democratic vote share is only positive and significant in the coefficient on the far left indicating the effect of mass shootings is highly sensitive.



APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Having such a low breakdown value suggests that any meaningful departure from smoothness, that is any departure of the slope changing between consecutive periods, would cause the observed effects in the GMAL data to not be significant. Similarly, the top right panel shows a sensitivity analysis for different values of $\overline{M}$ under the $RM$ approach. With a very small breakdown value of around 0.0, there is even further evidence that the results are highly sensitive. In post-treatment trends were even 1/20th the size of those observed in the pre-treatment set, we would include a null effect in our confidence sets. Finally, the bottom panel shows the results from the $SDRM$ approach. These results too suggest a high degree of sensitivity to even very modest departures from the assumption that the possible non-linearities in the post-treatment difference in trends are bounded by the observed non-linearities in the pre-treatment difference in trends. In total, these results indicate the effect of shootings on vote shares is highly sensitive to violations of the parallel-trends assumption.

We note one final thing about addressing violations of parallel trends. In choosing the method to address this core issue, it is important to consider the exact nature of the data one has available to them, the statistical power that they have, and the amount of corresponding numerical degrees of freedom. We are *not* arguing that every case should employ unit-specific trends, for example. What we *are* arguing is that all researchers should diagnose and address potential violations of this core assumption. *How* they do so—with the many tools at their disposal that we have outlined above—is less important than *that* they do so.

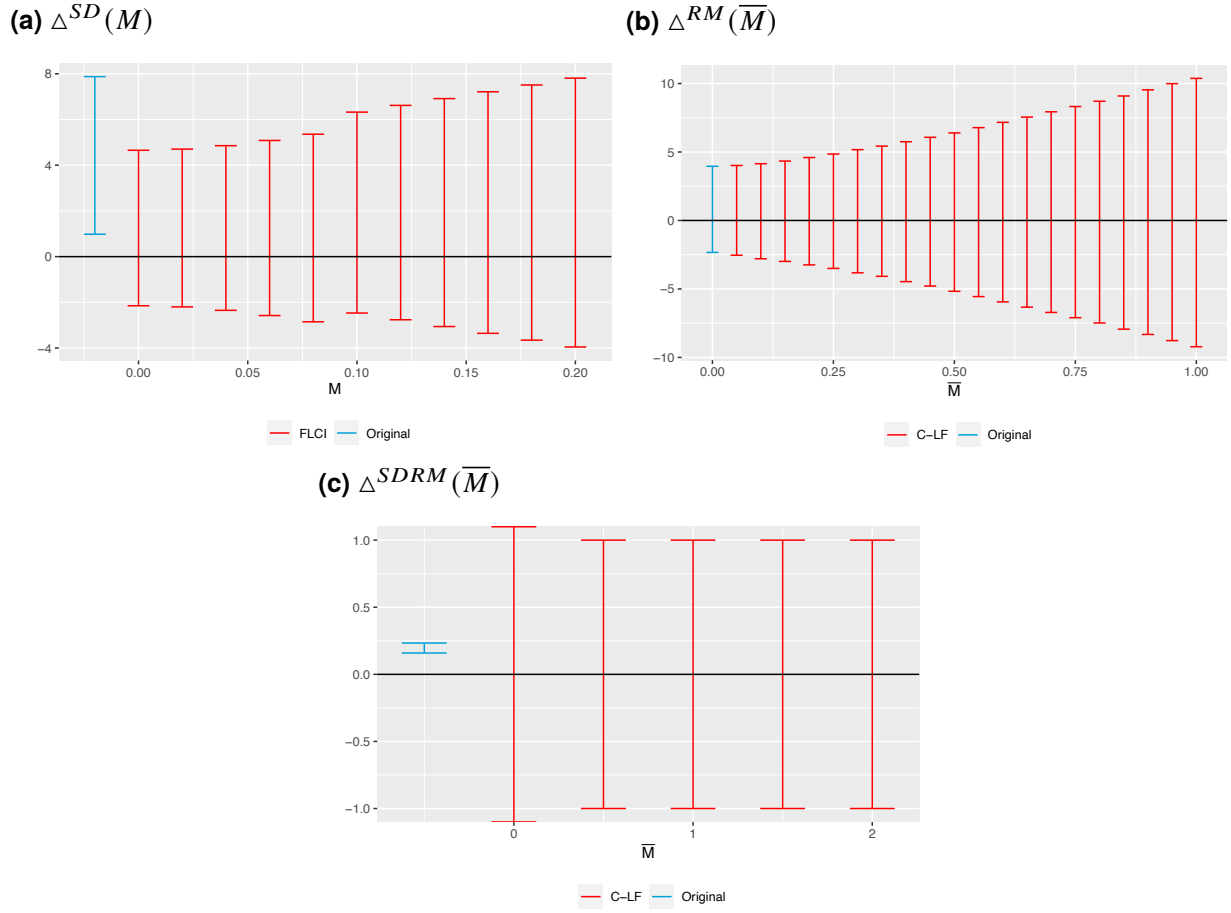
## 7. DIAGNOSING AND ADDRESSING TREATMENT EFFECT HETEROGENEITY/REMOVING CONTAMINATED COMPARISONS

Recent research has also shown that issues arise with the TWFE with variations in the treatment timing when there is heterogeneity of treatment effects across the time since treatment or across units. If there is heterogeneity in time since treatment only, the TWFE "corresponds with a potentially non-convex weighted average of the parameters" (Roth et al. 2022, 12). Goodman-Bacon (2021) shows that in this scenario, the TWFE can be written as "a convex weighted average of differences-in-differences comparisons between pairs of units and time periods in which one unit changed its treatment status and the other did not. Counterintuitively, however, this decomposition includes difference-in-differences that use as a 'control' group units who were treated in earlier periods. Hence, an early-treated unit can



APSR Submission Template

**FIGURE 8. Implementing Rambachan and Roth's Sensitivity Analysis in the Shooting Example**




Results from the sensitivity analysis suggested by Rambachan and Roth (2021) using GMAL's data; that is, testing for effect sensitivity across $\triangle^{SD}(M)$, $\triangle^{RM}(\overline{M})$, and $\triangle^{SDRM}(\overline{M})$. The models incorporate information from 3 elections prior to treatment and 5 post-treatment periods. **Takeaway:** The results show that the effect of shootings on vote shares is highly sensitive and does not hold with any meaningful deviation from parallel-trends.

get negative weights if it appears as a 'control' for many later-treated units" (Roth et al. 2022, 12).[40] The event study design does not address this concern; indeed, as Sun and Abraham (2021) show, using event-study regressions like the one in equation 5 may lead to erroneous conclusions in the presence of heterogeneity.

Goodman-Bacon (2021) provides an approach to decompose the 2x2 difference-in-difference estimates embedded in the TWFE with staggered treatment using the `bacondecomp` package in *R* and *STATA* Goodman-Bacon et al. (2019); Goodman-Bacon (2021).[41] We provide this decomposition in Figure 9.[42] (The weights for the Goodman-Bacon decomposition are reported in Tables S33 and

[40]For more details see Goodman-Bacon (2021) and Roth et al. (2022, 11-13).

[41]With this approach, we make the panel balanced and code all post-treatment units as treated.

[42]For another weighting decomposition approach, see de Chaisemartin et al. (2019). We provide a summary of

S34 in the Appendix.) As can be seen, a few interesting patterns emerge. First, in both the GMAL and HHB data the TWFE is a composite of 2x2's that elicit large negative and large positive effects. Depending on which of the 2x2's one includes in the estimate, the effect can be very different (as seen by looking at the spread of estimates across the y-axis). Second, it appears that the TWFE in this case are a composite that weights highly by several few comparisons of always treated vs. timing. However, many of the 2x2 estimates have similar weight—as noted by the cluster of estimates on the left side of the graph. Overall, to our eye, there appears to be no clear evidence our effects are driven by treatments of various types. Still, we think it important to note that for several reasons the mass shooting example is not an ideal application to show the value of Goodman-Bacon decomposition as (at present) the Goodman-Bacon decomposition only decomposes the TWFE and does not decompose the more sophisticated models we implemented to account for potential parallel-trends assumption violations. Nevertheless, what we observe in the mass shootings context may not always be true, and so scholars using difference-in-differences estimators with staggered treatment should adopt this decomposition as a standard diagnostic test to illuminate the extent to which the TWFE is driven by specific types of comparisons.

**FIGURE 9. Illustration of the Goodman-Bacon Decomposition of the TWFE Models**

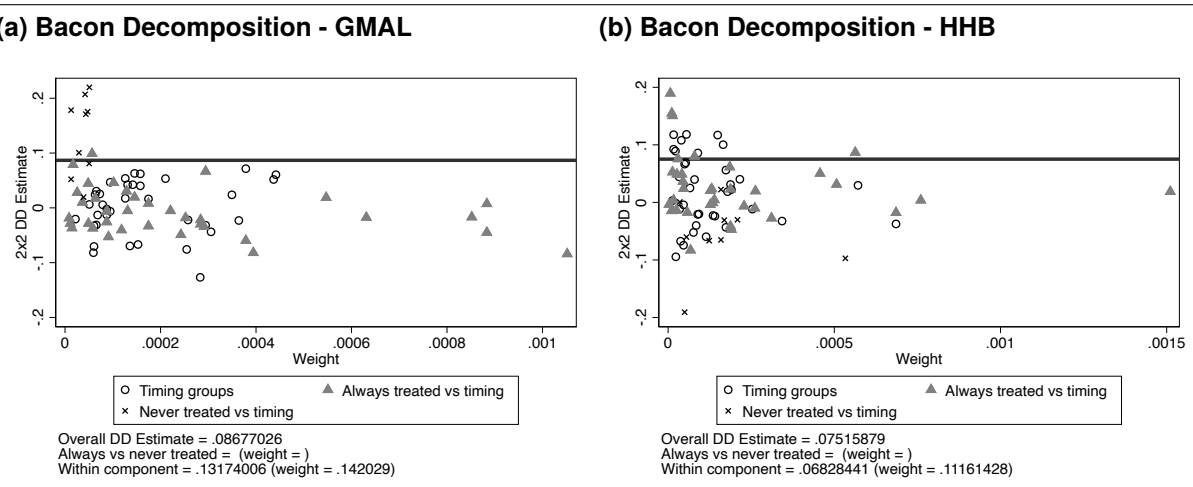



Figure shows the results from the Bacon decomposition for the TWFE models. The figure shows all of the possible 2x2 difference-in-differences, with their weights for the ATE on the x-axis and the effect size on the y-axis. **Takeaway:** There are no comparisons with negative weights in our application of the effects of mass shootings on electoral outcomes and no clear evidence effects are driven by treatments of various types.

these weights in Tables S8 and S9 in the Appendix.

Several solutions to the problems that arise with heterogeneous treatment effects have been proposed.[43] Some allow for the extension of their approach to include additional covariates beyond two-way fixed effects, including unit-specifc trends; others do not. One approach that does allow for the extension

[43]Three other solutions to treatment effect heterogeneity problems identified in the literature are worth mentioning. The differences between these are nuanced and not all may be well-suited in some applications. First, like Sun and Abraham (2020), Callaway and Sant'Anna (2021) argue that scholars should use a method that restricts to "clean comparisons" applying to scenarios where "(i) multiple time periods, (ii) variation in treatment timing, and (iii) when the 'parallel-trends assumption' holds potentially only after conditioning on observed covariates." This approach facilitates the estimation of propensity scores conditional on observed covariates to help achieve pre-treatment balance. With this approach, we make the panel balanced and code all post-treatment units as treated as doing so is more appropriate for this approach. We show this approach in Figure S19 in the Online Appendix. Unfortunately, this approach has limited value in our application for two reasons. First, *even when* one uses "clean comparions" as suggested by Callaway and Sant'Anna (2021) and covariates, differential pre-treatment trends issues remain. Second, their approach does not yet extend to models with unit-specific time trends. These may be less of an issue in other applications, so we include these as an illustration of this method and its results. Second, De Chaisemartin and d'Haultfoeuille (2020) provide an alternate approach for assessing and addressing implemented in the `did_multiplegt` package in *STATA* and `DIDmultiplegt` package in *R* de Chaisemartin et al. (2019); Zhang (2022). With this approach, we code all post-treatment units as treated as doing so is more appropriate for this approach. Unfortunately, this approach only allows linear trends and doesn't allow for flexibility in other model parameters that approaches like Sun and Abraham (2021) afford and is more computationally intensive. Still, to illustrate this method, we provide the results for this in Tables S8 and S9 in the Appendix. Finally, (Borusyak et al. 2021, 1) use "an intuitive 'imputation' form [where] treatment-effect heterogeneity is unrestricted." This approach is implemented in the `did_imputation` package in *STATA* and `didimputation` package in *R* (Borusyak 2022; Butts 2022). We use the treatment of coding treatment only in the current period as it is more appropriate to do so for this approach. When we implement this approach, we still see a sizable effect both pre- and post-treatment in the TWFE. Unfortunately, this approach is not currently designed to implement with unit specific trends in our example. Though the package technically does allow trends, the help file warns users to "Use [trends] with caution: the command may not recognize that imputation is not possible for some treated observations." This appears to be the case in our application.



APSR Submission Template

of including unit-specific trends is Sun and Abraham (2021).[44] This approach is implemented in the `eventstudyinteract` package in *STATA* and `fixest` package in *R* (Sun 2021; Berge et al. 2022).[45] This approach "estimates the shares of cohort as weights." In our case, implementing Sun and Abraham's solution with with a simple TWFE significantly cuts down on the effect estimates provided by GMAL. At first, these changes look modest. In the first treatment period, the event study estimates go from 3 percentage points (p = 0.000) in the naive models to 2.4 percentage points (p = 0.001) in the Sun and Abraham adjusted models. However, in the second and following treatment periods, the effect that is large as 10-13 percentage points in the naive event study heavily attenuates and even becomes negative (though not significant—being 1.1 (p=0.19), -0.4 (p=0.75), -2.2 (p=0.12), and -1.5 (p=0.18) percentage points in post-treatment elections 1-4 respectively. Once we add unit-specific trends to Sun And Abraham's estimator, the effect gets even smaller. We go from 2.4 percentage points (p = 0.001) in the TWFE to 1.46 percentage points in the model with linear trends (p=0.031), quadratic (p=0.051), or cubic trends (p=0.085)—with the results becoming less significant with each. (Even in the cubic model, the standard error remains modest in size–being 0.8 percentage points.) With trends, the long-run effect of 10-13 percentage points is not present. Moreover, none of the effects are present in the HHB data. This suggests that effect heterogeneity plays some role. Once adjusted for, long-term effects attenuate substantially and short-term effects become much more modest and flimsier to reasonable specifications within the realm of researcher's arbitrary decision-making (e.g. the coding of treatment, the functional form of the unit-specific trends that one includes, and election differences between HHB and GMAL).

For reasons we outline below, we think it unwise to cherry-pick one model specification above. Combining the evidence from all of the various approaches taking into account potential contamination from treatment effect heterogeneity, the best evidence suggests that 1.) violations of parallel trends loom large in this context, 2.) effect heterogeneity may play a modest role in this application, and 3.) there is no sign of the sizable and durable effect on Democratic vote shares, but perhaps a much smaller effect—and one that is not clearly distinct from zero across all reasonable model specifications and is sometimes even negative in some specifications. That said, these lessons may not always be true in

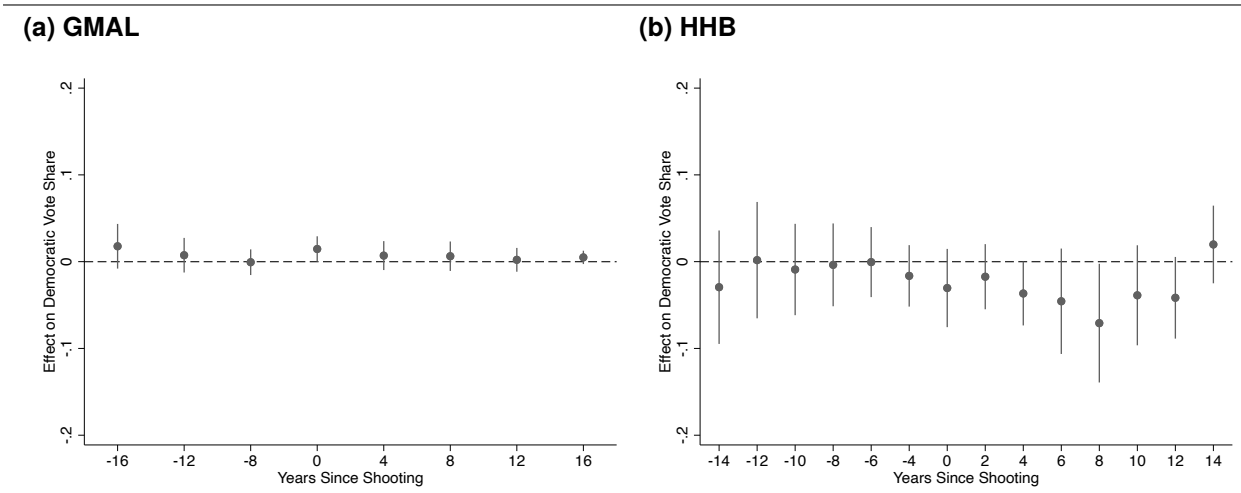
[44]Our results are robust to excluding never-treated observations using Sun and Abraham (2021).

[45]With this approach we make the panel strongly balanced and use the treatment in the current period coding.



APSR Submission Template

**FIGURE 10. Sun and Abraham's (2020) Approach For Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects**




Results from the clean comparisons suggested by Sun and Abraham (2021) using GMAL and HHB data. We focus on these two for two reasons 1.) these treatments are most similar 2.) the Yousaf panel and coding of mass shootings does not recover pre-treatment balances using Sun and Abraham (2021)'s method. Models include quadratic county-specific time trends to address potential violations of the parallel-trends assumption in the TWFE. **Takeaway:** Clean comparison effects with trends show no sign of a sizable and durable effect on Democratic vote shares shown in the TWFE nor in the simple event-study plot (see Figure 5 above).

other applied contexts. Hence, we recommend that scholars include tests for both parallel trends and treatment heterogeneity and implement the solutions suggested in the literature as we have done above.

## 8. ASSESSING OTHER POTENTIAL ISSUES RAISED IN THE ECONOMETRICS LITERATURE ON EXECUTING A DIFFERENCE-IN-DIFFERENCES DESIGN

A few final words of guidance and caution remain when using the difference-in-differences approach. First, in all of the empirical checks, it's important to not forget theory. For example, in the analyses above, we have focused on whether shootings change Democratic vote shares in the counties in which they occur. We have done so because whether or not an effect occurs here is the central dispute in this literature. However, it is possible this is not conceptually how mass shootings' effects work. Perhaps shootings have spillover effects—with any effects showing up in adjacent counties—or effects of shootings arise as a function of the distance to where a shooting occurs, the time since a shooting occurred, or the intensity of the shooting itself (e.g., the number of deaths/injuries in a shooting). Alternatively, perhaps mass shootings have effects that show up at a national level. The best evidence we



APSR Submission Template

have currently suggests that none of these things occur in the mass shootings contexts (HHB 2020).[46] However, in their analyses, scholars should not forget there are often multiple ways of conceptualizing treatment exposure.

Second, we have shown the importance of including unit-specific time trends to ensure that the parallel-trends assumption is not violated. However, it is possible that other trends may bias the effect of interest. For example, nationwide trends in time-varying covariates may drive results.[47] Including these might change results. For example, using GMAL data, if we include the covariates they do (population, proportion non-white, and change in the unemployment rate), the effect estimate for a model with county-specific time trends is 0.7 percentage points ($p<0.133$; 95% CI: [-0.2, 1.6]).[48] However, adding these controls' linear interaction with time, the effect estimate is even smaller—being only 0.2 percentage points—with less evidence for a meaningful effect ($p<0.597$; 95% CI: [-0.7, 1.2]). This may be a useful check for scholars of difference-in-differences designs to include.

Third, in some instances it may be useful to unpack the treatment effects at a more granular level—estimating, for example, the effect of individual shootings, rather than the average effect. Scholars have developed the synthetic control method for such instances (Abadie et al. 2015; Kreif et al. 2016; Porreca 2022; Arkhangelsky et al. 2021) providing a systematic way to choose counterfactual units when individual treated units are of interest. Similarly, if one is interested in whether a small set of observations drives results, Broderick et al. (2020) have developed an effective and computationally feasible procedure[49] package in `R`..

Fourth, scholars may be interested in estimating distributional effects which has advanced rapidly in recent years. Recent work has combined difference-in-differences estimators with those produced

[46] Only HHB considers these different types of coding treatment. They find null effects with all of these treatments once time trends are accounted for.

[47] More generally, when including covariates it is important that the controls one uses do not include those that are affected by treatment. Otherwise, these would be "bad controls" (Montgomery et al. 2018). This does not appear to be the case in the mass shootings context; after all, only GMAL use (a few) controls (i.e. population, proportion nonwhite, and change in unemployment rate) and these do not influence their substantive results. But researchers should keep this fact in mind when choosing controls.

[48] From the models with quadratic county-specific time trends.

[49] See their `zaminfluence`

from quantile regression (see Roth et al. 2022; Callaway and Li 2019). Using this test, HHB show little signs of shifts at any point along the distribution of Democratic vote shares; that is, there is little evidence for shootings sparking polarization of electoral outcomes (see their Figure A12 and surrounding discussion).

Finally, in making modeling decisions in the difference-in-differences space, one needs to acknowledge that there may be tradeoffs between bias and precision. For example, using higher order polynomials for unit specific time trends requires more power and may inflate standard errors. For this reason, in our applied example, we have taken great care to pay attention to both effect sizes and statistical significance, and the range of potential effects. Recent work has shown the importance of considering power in testing for pre-trends (Roth 2022; Freyaldenhoven et al. 2019; Roth et al. 2022).

In sum, we note that, to a certain extent, how to implement difference-in-differences designs depends on the nature of the data —that is, whether or not there are likely violations of the parallel-trends assumption and/or unaccounted treatment effect heterogeneity. In examining the effect of gun violence on electoral outcomes, the former appears to be the key issue to identification while the latter is less of an issue. However, this may not always be the case. As such, we think it best for authors to follow the suggestions we outline above to ensure their inferences are not misleading. In our applied example, doing so reconciles why different studies using the same data have come to vastly different conclusions about the effects of gun violence on electoral vote shares.

## 9. SYNTHESIZING THE EVIDENCE FROM MULTIPLE DIFFERENCE-IN-DIFFERENCES SPECIFICATIONS

Before concluding, we believe it important to discuss how researchers should interpret their results when there are many different model specifications, as with difference-in-differences designs. A full discussion of synthesizing multiple model results is beyond the scope of this paper. However, we note briefly here a few important points required to come to a conclusion about the effects of mass shootings on election outcomes.

In our particular example, there are a multitude of plausible models that might be run, a small number of which show statistically significant effects (both positive and negative). Given the potential for bias and the role that researcher degrees of freedom play (even unintentionally), we think it important for



APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

researchers to 1.) address the potential threats to inference we outline above, 2.) be transparent about the role that simple changes to model specification play, and 3.) take a "preponderance of evidence" rather than a "singular model" approach.

It is *essential* that researchers running difference-in-differences model specifications do not cherry-pick individual model specifications, but rather test for robustness across the dimensions discussed above. Increasing the number of specifications run makes it potentially easier to choose one model for incorrect reasons (e.g. choosing a specification just because it achieves statistical significance or being led by unintentional internal biases to justify a preferred model).[50]

What does this mean in the mass shootings context? Though on occasion we see intermittent statistically significant effects, these effects are 1.) much smaller than previous research has suggested and 2.) not robust to reasonable changes to these specifications under the control of researchers. Taking a singular model approach, we become vulnerable to the curse of researcher degrees of freedom and mistakenly conclude that mass shootings have an effect (either positive or negative). However, with a preponderance of evidence approach, we find strong reasons to doubt mass shootings have significant, systematic, or large effects on Democratic vote shares.[51] Models that account for violations of parallel trends provide little to no evidence mass shootings cause large and meaningful electoral change in the United States and fairly compelling evidence that is consistent with a null effect.

This can be seen by synthesizing four pieces of evidence. First, though some corrected models show a much smaller, but perhaps very modest, positive effect on Democratic vote share, most of these estimates are not statistically significant. Second, negative effects show up fairly often across the small, but reasonable, changes to model specification well within the control of researchers and their many

[50] In comparing the various approaches outlined above, it may be useful to leverage new programming tools that make it easier to estimate multiple approaches at once. For one example, we point to the 2021 GitHub repository of Florian M. Hollenbach found here.

[51] Arguing that scholars should be running model specifications may prompt issues with multiple comparisons. We believe that scholars should be careful to not over-interpret singularly significant effects in a deluge of otherwise not significant results. However, properly adjusting for multiple comparisons across similar robustness checks is not well-developed. And we note here, in arguing more for the null, it is more conservative for us to not make any adjustments for multiple comparisons.

degrees of freedom. We can see this visually in Figure 11 which plots the distribution of effects and p-values for all of the difference-in-differences models and event-study models that we estimate above. As can be seen in the left top panel, all of the coefficients from models with trends are much smaller than those provided by GMAL's TWFE. Some are positive and some are negative; but the distribution spikes near zero. The average effect in the difference-in-differences model is 0.9 percentage points. As can be seen in the bottom left panel, in the event study models there is also a spike at zero, with a similar amount of positive and negative effects. The average effect for all post-treatment periods 0-4 across all event-study models is 0.4 percentage points and the average across all models in the year immediately after treatment (i.e. period 0) is 0.07 percentage points. Third, when significant and positive effects do show up these effects are often not robust to slight variations in model specification that are within the degrees of freedom researchers face. Fourth, sensitivity analyses that embrace the uncertainty around exact departures from parallel trends show that the results are *highly* sensitive to even minimal reasonable departures from parallel trends. Hence, the preponderance of evidence suggests that a large effect is implausible and modest positive effects are anything but sure.

## 10. CONCLUSION

In reconciling research on the effects of mass shootings on electoral outcomes, our work has also highlighted the considerations we argue should become standard practice given the potential hazards of navigating difference-in-difference designs. In addition to resolving an important question, we hope our article will spark a more nuanced approach to estimating difference-in-differences models—one more closely aligned with the methodological treatment of this oft-used promising research design. If appropriately used, the checks we have outlined above will help researchers make better inferences using this commonly used identification strategy.

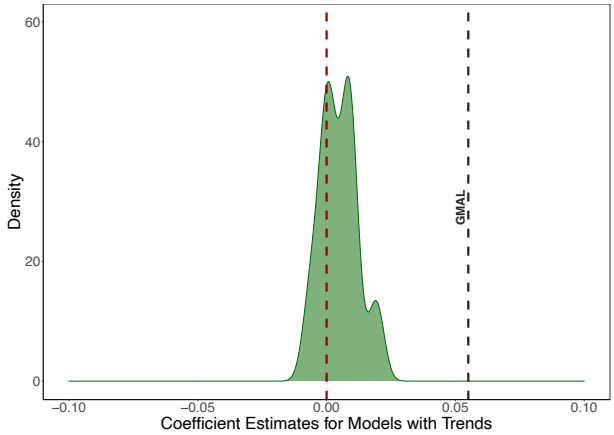
In this regard, our work does have some limitations. The methodological contribution we provide most readily applies to cases where the treatment one desires to estimate may not be fully exogenous, the treatment one desires to estimate is an event that may vary in timing across units, and there are a comparatively larger sample size with more cross-sections than time points. Instances that depart from these may leverage similar approaches that we outlined above, but may also have unique features. Moreover, we have not explored some aspects of panel data estimation that are more recently developed



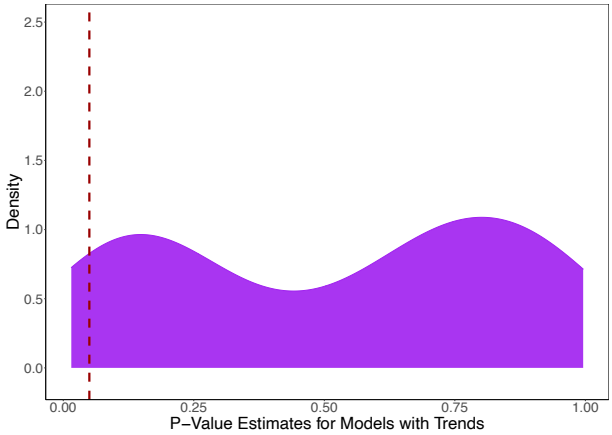
APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

**FIGURE 11. Distribution of All Effect Estimates and P-Values for Models with County Trends**

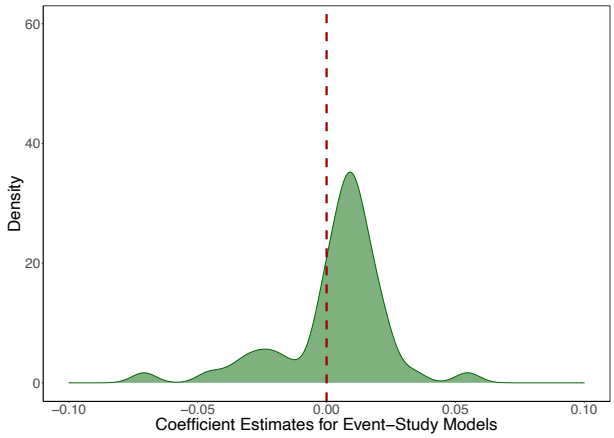
**(a) Distribution of Coefficients in Diff-in-Diffs Models with County Trends**



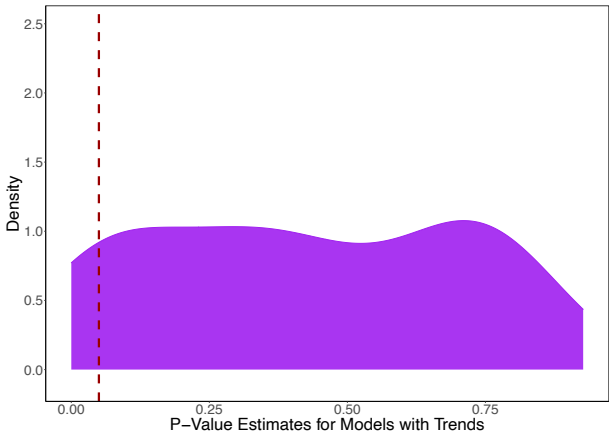

**(b) Distribution of P-Values in Diff-in-Diffs with County Trends**




**(c) Distribution of Coefficients in Event-Study Models with County Trends**




**(d) Distribution of P-Values in Event-Study Models with County Trends**




Distribution of all model estimates with trends in Figure 5 above in the first row and then for all the event study estimates in the paper on the bottom row. The event study coefficients are shown for periods 0-4 post treatment. The left panel in each row shows coefficients (in percentage point units). The right panel in each row shows the distribution of p-values across model specifications. **Takeaway:** Once we account for potential violationsof parallel trends, the effects of shootings spike at zero, are only rarely significant, are not robust to slight changes in model specification, and sometimes positive and sometimes negative.

for scenarios with very few treatment units (e.g. synthetic controls) that are valuable, but beyond the scope of what we do here. Finally, our work is applied to a context where there is not currently, nor any prospect of a future, experimental baseline. While the econometric literature we draw from has a long history of highlighting the value of the checks we run—by using proofs, simulations, and other validation techniques—there has yet to be (to our knowledge) a comparison of the diff-in-diffs tools in our arsenal to a randomized baseline. Future work would do well to find other contexts where



APSR Submission Template

randomization is possible and add this benchmarking task to our suite of studies on this widely used method, as has been done with other methodological techniques (e.g., Arceneaux et al. 2006; Green et al. 2009).

Returning to the context of this study, America's legacy of gun violence is heartbreaking and the thousands of deaths that occur from guns each year constitute a policy failure of epic proportions. Yet, whether policymakers relative inaction occurs in spite of (or as a result of a lack of) an electoral response has been an area of disagreement in previous work. While agreeing that mass shootings do not effect voter turnout, scholars have come to vastly different conclusions about the effect of mass shootings on vote shares. In this paper, we show that gun violence has little to no impact on vote shares and that the previous work that has shown a relationship failed to navigate the many pitfalls that come with difference-in-differences designs.

Taken together, these checks show we cannot support the hypothesis that suggest that mass shootings (or school shootings or 'rampage-style' school shootings) increase Democratic vote shares substantially. Moreover, even the most generous interpretation—i.e. one that ignores statistical uncertainty around the estimates altogether (something that we think is neither wise or nor prudent)—suggests that shootings have, if anything, a very modest effect on Democratic vote share—one that is *much* smaller than suggested by prior research. When looking across all robustness checks, we cannot conclude that mass shootings—be they in schools or not, or rampage-style or not—substantially affect election outcomes. Such a conclusion comes from results that are not robust and that are highly sensitive.

Furthermore, we also note that these estimates are all *local* to the county in which the shooting occurred. Though the presence of any mass shootings is, in our view, repugnant, mass shootings are (thankfully) a relatively rare phenomena. Given that shootings only occurred in 0.4% of counties (116 total; 11.6 per election in the sample) in the HHB dataset, 0.4% of counties (115 total; 11.5 per election in the sample) in the GMAL dataset, and only 0.5% (72 total; 14.4 per election in the sample) of counties in the Yousaf data further emphasizes the limited impact shootings have had on elections.[52] Putting these two pieces together—both the modest effect sizes in percentage points and their limited scope—suggests effects of mass shootings are of little substantive consequence for election outcomes.

[52]Only HHB consider whether there are spillover effects on adjacent counties, and find none (p. 1377).



APSR Submission Template

*Even if* we take the point estimates above at face value and ignore statistical uncertainty (something we think we should *not* do), a county-specific effect of the size we observe and the infrequency at which they happen would have *virtually no effect on any statewide or national election*.

Scholars doing work on the localized electoral effects of plausibly exogenous events or shocks should not forget there are often multiple ways of conceptualizing treatment exposure. These include, but are not limited to, treatments that conceptualize treatment as being short-lived (i.e. constrained only to the period when they happen) or longer-term (i.e. turned on in all periods after treatment occurs), treatments that consider spillover effects on units adjacent to treatment, treatments that consider the dosage or intensity of treatment, and treatments that consider the possibility of national treatments drowning out any potential local effects. The best evidence we have currently suggests that none of these things occur in the mass shootings contexts (see Hassell et al. (2020, p. 1377) for more details).

Our work sets the table for future work on the political economy of gun violence and retrospective voting/accountability more generally.[53] The lingering question is why mass shootings fail to substantively change the electoral incentives elected officials face. Despite having favorable conditions for a response, little to no detectable retrospective voting occurs as a result of mass shootings. This result illuminates a need for a broader research agenda that better explores the nature of retrospective voting (or lack thereof). Our paper provides important context by showing that some of the most common characteristics thought to promote active retrospective voting (e.g. voter attention, media coverage, potential for governmental action) are not always sufficient to spark electoral change, providing answers to important discrepancies in previous work on the effect of mass shootings on electoral outcomes and a guide for researchers attempting to navigate the potential pitfalls of difference-in-difference designs.

[53] Future research should consider *why* mass shootings fail to mobilize new voters nor change the voting patterns of existing voters.



APSR Submission Template

## REFERENCES

Abadie, Alberto , Susan Athey, Guido W Imbens, and Jeffrey Wooldridge (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Abadie, Alberto , Alexis Diamond, and Jens Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science 59*(2), 495–510.

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics*.

Angrist, Joshua D and Jörn-Steffen Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Arceneaux, Kevin , Alan S Gerber, and Donald P Green (2006). Comparing experimental and matching methods using a large-scale voter mobilization experiment. *Political Analysis 14*(1), 37–62.

Arkhangelsky, Dmitry , Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager (2021). Synthetic difference-in-differences. *American Economic Review 111*(12), 4088–4118.

Armitage, Seth (1995). Event study methods and evidence on their performance. *Journal of Economic Surveys 9*(1), 25–52.

Bai, Jushan (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, Jushan and Serena Ng (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association 116*(536), 1746–1763.

Baker, Andrew C , David F Larcker, and Charles CY Wang (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics 144*(2), 370–395.

Barney, David J and Brian F Schaffner (2019). Reexamining the effect of mass shootings on public support for gun control. *British Journal of Political Science 49*(4), 1555–1565.

Beck, Nathaniel L , Jonathan N Katz, and Umberto G Mignozzetti (2014). Of nickell bias and its cures: Comment on gaibulloev, sandler, and sul. *Political Analysis 22*(2), 274–278.

Berge, Laurent , Sebastian Krantz, and Grant McDermott (2022). fixest: Fast fixed-effects estimations. *CRAN*.

Bilinski, Alyssa and Laura A Hatfield (2018). Nothing to see here? non-inferiority approaches to parallel trends and other model assumptions. *arXiv preprint arXiv:1805.03273*.

Binder, John (1998). The event study methodology since 1969. *Review of Quantitative Finance and Accounting 11*(2), 111–137.

Borusyak, Kirill (2022). Did_imputation: Stata module to perform treatment effect estimation and pre-trend testing in event studies.

Borusyak, Kirill , Xavier Jaravel, and Jann Spiess (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv Preprint* (2108.12419).

Broderick, Tamara , Ryan Giordano, and Rachael Meager (2020). An automatic finite-sample robustness metric: Can dropping a little data change conclusions? *arXiv Preprint* (2011.14999).

Butts, Kyle (2022). didimputation: Imputation estimator from borusyak, jaravel, and spiess (2021). *CRAN*.



APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Callaway, Brantly and Tong Li (2019). Quantile treatment effects in difference in differences models with panel data. *Quantitative Economics 10*(4), 1579–1618.

Callaway, Brantly and Pedro HC Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Cameron, A Colin and Douglas L Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Clarke, Damian and Kathya Tapia-Schythe (2021). Implementing the panel event study. *The Stata Journal 21*(4), 853–884.

De Chaisemartin, Clément and Xavier d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

De Chaisemartin, Clément and Xavier D'Haultfoeuille (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. Technical report, National Bureau of Economic Research.

de Chaisemartin, Clément , Xavier D'Haultfoeuille, and Yannick Guyonvarch (2019). Did_multiplegt: Stata module to estimate sharp difference-in-difference designs with multiple groups and periods.

DeSilver, Drew (2016). The growing democratic domination of nation's largest counties. *Pew Research Center*.

Freyaldenhoven, Simon , Christian Hansen, Jorge Pérez Pérez, and Jesse M Shapiro (2021). Visualization, identification, and estimation in the linear panel event-study design. Technical report, National Bureau of Economic Research.

Freyaldenhoven, Simon , Christian Hansen, Jorge Pérez Pérez, and Jesse Shapiro (2022). Xtevent: Stata module to estimate and visualize linear panel event-study models.

Freyaldenhoven, Simon , Christian Hansen, and Jesse M Shapiro (2019). Pre-event trends in the panel event-study design. *American Economic Review 109*(9), 3307–38.

Garcia-Montoya, Laura , Ana Arjona, and Matthew Lacombe (2022). Violence and voting in the united states: How school shootings affect elections. *American Political Science Review 116*(3), 807–826.

Gelman, Andrew and Eric Loken (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University 348*.

Gobillon, Laurent and Thierry Magnac (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics 98*(3), 535–551.

Goodman-Bacon, Andrew (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Goodman-Bacon, Andrew , Thomas Goldring, and Austin Nichols (2019). Bacondecomp: Stata module to perform a bacon decomposition of difference-in-differences estimation.

Green, Donald P , Terence Y Leong, Holger L Kern, Alan S Gerber, and Christopher W Larimer (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis 17*(4), 400–417.



APSR Submission Template

Grimmer, Justin , Eitan Hersh, Marc Meredith, Jonathan Mummolo, and Clayton Nall (2018). Obstacles to estimating voter id laws' effect on turnout. *Journal of Politics 80*(3), 1045–1051.

Hansen, Ben B and Jake Bowers (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 219–236.

Hartman, Erin and F Daniel Hidalgo (2018). An equivalence approach to balance and placebo tests. *American Journal of Political Science 62*(4), 1000–1013.

Hartman, Todd K and Benjamin J Newman (2019). Accounting for pre-treatment exposure in panel data: Re-estimating the effect of mass public shootings. *British Journal of Political Science 49*(4), 1567–1576.

Hassell, Hans JG , John B Holbein, and Matthew Baldwin (2020). Mobilize for our lives? school shootings and democratic accountability in us elections. *American Political Science Review 114*(4), 1375–1385.

Healy, Andrew and Gabriel S. Lenz (2017). Presidential voting and the local economy: Evidence from two population-based data sets. *Journal of Politics 79*(4), 1419–1432.

Kahn-Lang, Ariella and Kevin Lang (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics 38*(3), 613–620.

Kreif, Noémi , Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics 25*(12), 1514–1528.

Liu, Licheng , Ye Wang, and Yiqing Xu (2021). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *arXiv Preprint* (2107.00856).

Manski, Charles F and John V Pepper (2018). How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions. *Review of Economics and Statistics 100*(2), 232–244.

Marcus, Michelle and Pedro HC Sant'Anna (2021). The role of parallel trends in event study settings: an application to environmental economics. *Journal of the Association of Environmental and Resource Economists 8*(2), 235–275.

Marsh, Wayde Z.C. (2022). Trauma and Turnout: The Political Consequences of Traumatic Events. *American Political Science Review*.

Montgomery, Jacob M , Brendan Nyhan, and Michelle Torres (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science 62*(3), 760–775.

Montiel Olea, José Luis and Mikkel Plagborg-Møller (2019). Simultaneous confidence bands: Theory, implementation, and an application to svars. *Journal of Applied Econometrics 34*(1), 1–17.

Mou, Hongyu , Licheng Liu, and Yiqing Xu (2022a). Package 'panelview'.

Mou, Hongyu , Licheng Liu, and Yiqing Xu (2022b). panelview: Panel data visualization in r and stata. *Available at SSRN 4202154*.

Musu-Gillette, Lauren , Anlan Zhang, Ke Wang, Jana Kemp, Melissa Diliberti, and Barbara A. Oudekerk (2018). Indicators of school crime and safety: 2017. *National Center for Educational Statistics* (NCES 2018-036).

Porreca, Zachary (2022). Synthetic difference-in-differences estimation with staggered treatment timing. *Available at SSRN*.



APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template APSR Submission Template

Rambachan, Ashesh and Jonathan Roth (2021). An honest approach to parallel trends. *Unpublished manuscript, Harvard University*.

Rogowski, Jon C and Patrick D Tucker (2019). Critical events and attitude change: Support for gun control after mass shootings. *Political Science Research and Methods 7*(4), 903–911.

Rossin-Slater, Maya , Molly Schnell, Hannes Schwandt, Sam Trejo, and Lindsey Uniat (2020). Local exposure to school shootings and youth antidepressant use. *Proceedings of the National Academy of Sciences 117*(38), 23484–23489.

Roth, Jonathan (2022). Pre-test with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*.

Roth, Jonathan , Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe (2022). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *arXiv Preprint 2201.01194*.

Schmidheiny, Kurt and Sebastian Siegloch (2019). On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *Equivalence, and Generalization (January 2019)*.

Sides, John , Lynn Vavreck, and Christopher Warshaw (2022). The effect of television advertising in united states elections. *American Political Science Review 116*(2), 702–718.

Sun, Liyang (2021). Eventstudyinteract: Stata module to implement the interaction weighted estimator for an event study.

Sun, Liyang and Sarah Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

U.S. Government Accountability Office (2020). K-12 education: Characteristics of school shootings. *GAO-20-455*.

Wing, Coady , Kosali Simon, and Ricardo A Bello-Gomez (2018). Designing difference in difference studies: best practices for public health policy research. *Annual review of Public Health 39*.

Xu, Yiqing (2022). Causal inference with time-series cross-sectional data: A reflection. *Available at SSRN 3979613*.

Yousaf, Hasin (2021). Sticking to one's guns: Mass shootings and the political economy of gun control in the united states. *Journal of the European Economic Association*.

Zhang, Shuo (2022). Didmultiplegt: Estimation in did with multiple groups and periods. *CRAN*.



APSR Submission Template

# Navigating Potential Pitfalls in Difference-in-Differences Designs: Reconciling Conflicting Findings on Mass Shootings' Effect on Electoral Outcomes - Online Appendix

# List of Tables

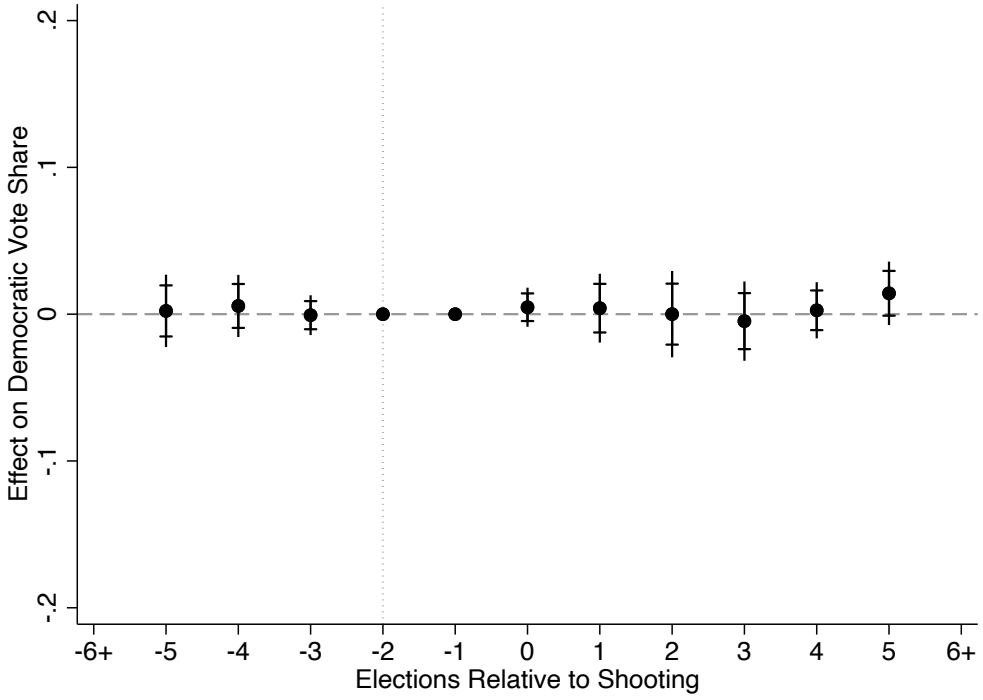
# List of Figures

Table S1: Differences in All Studies on the Effects of Gun Violence on Electoral Vote Shares

| | | **GMAL** | **Yousaf** | **HHB** |
|---|---|---|---|---|
| *Data* | Shootings | "Rampage-style" school shootings: "Rampage-style" shootings are shootings that "take place on a school-related public stage before an audience; involve multiple victims, some of whom are shot simply for their symbolic significance or at random; involve one or more shooters who are students or former students of the school and where the motivation of the shooting [does not] correlate with gang violence or targeted militant or terroristic activity" (GLAM, 1) | Mass Shootings: Mass shootings are all shootings "leading to four or more deaths at one location" (Yousaf, 2770) | All school shootings (HHB, 1377) |
| | Years | 1980 to 2016 | 2000 to 2016 | 2000 to 2018 |
| | Vote Outcomes | Presidential election returns only | Presidential, gubernatorial, senatorial, and congressional election returns from presidential election years only | Presidential, congressional, state, and local election returns in all years |
| *Methods* | Model Specifications | Difference-in-Differences TWFE **(No county specific time-trends included)** | Difference-in-Differences TWFE **(No county specific time-trends included)** | Difference-in-Differences TWFE with county specific time-trends |
| | Standard Errors | Clustered at the state level | Clustered at the state level | Clustered at the county (treatment) level |

Figure S1: Results Alternate Baseline Periods in Event Study Design that Accounts for County Specific Time Trends

(a) LINEAR -2 Period Benchmark




(b) QUADRATIC -2 Period Benchmark

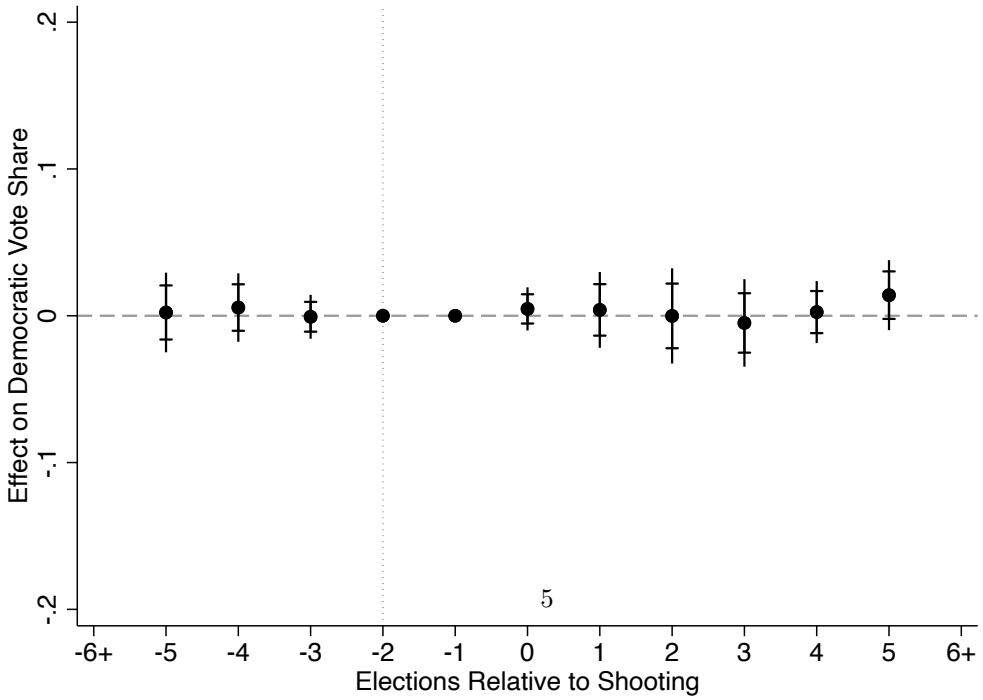
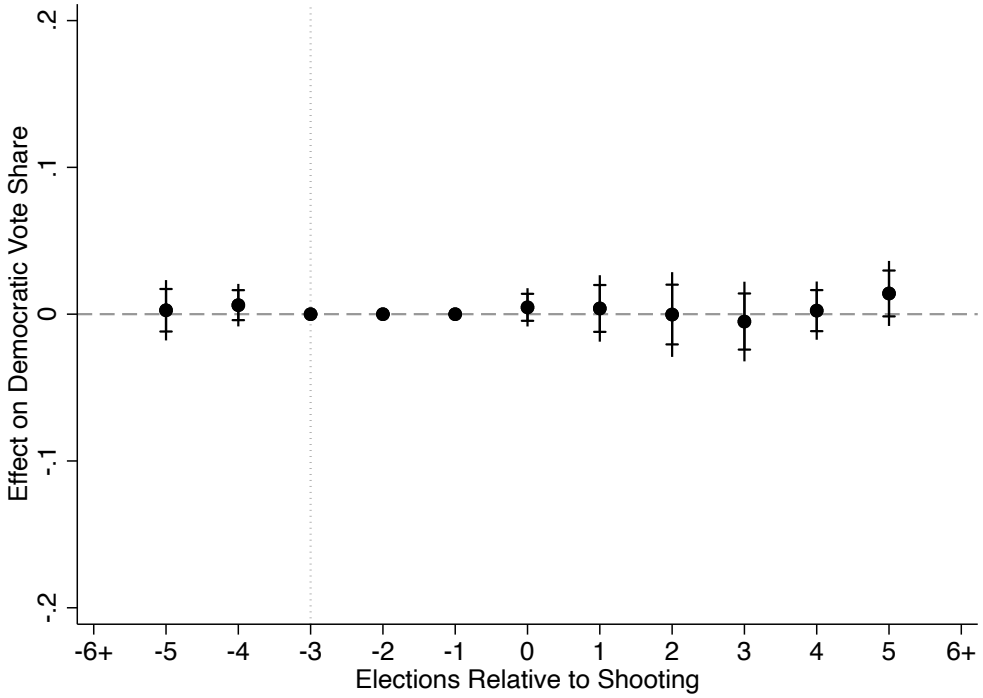





Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021). **Takeaway:** Benchmarked to pre-treatment trends at t-2, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

Figure S2: Results Alternate Baseline Periods in Event Study Design that Accounts for County Specific Time Trends (cont'd)

(a) LINEAR -3 Period Benchmark




(b) QUADRATIC -3 Period Benchmark

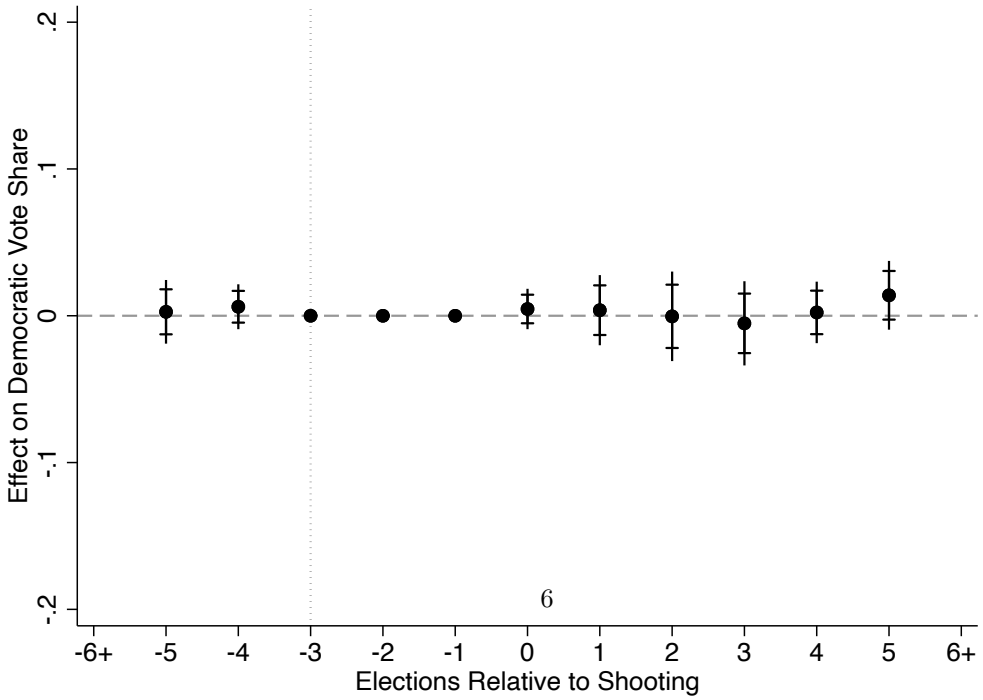
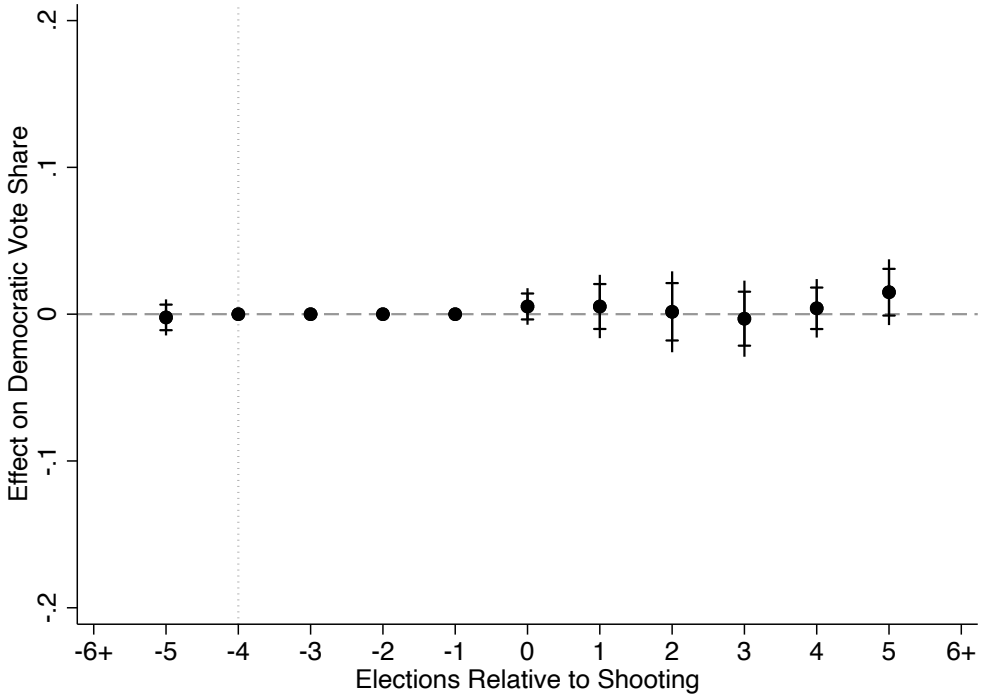





Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021). **Takeaway:** Benchmarked to pre-treatment trends at t-3, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

Figure S3: Results Alternate Baseline Periods in Event Study Design that Accounts for County Specific Time Trends (cont'd)

(a) LINEAR -4 Period Benchmark




(b) QUADRATIC -4 Period Benchmark

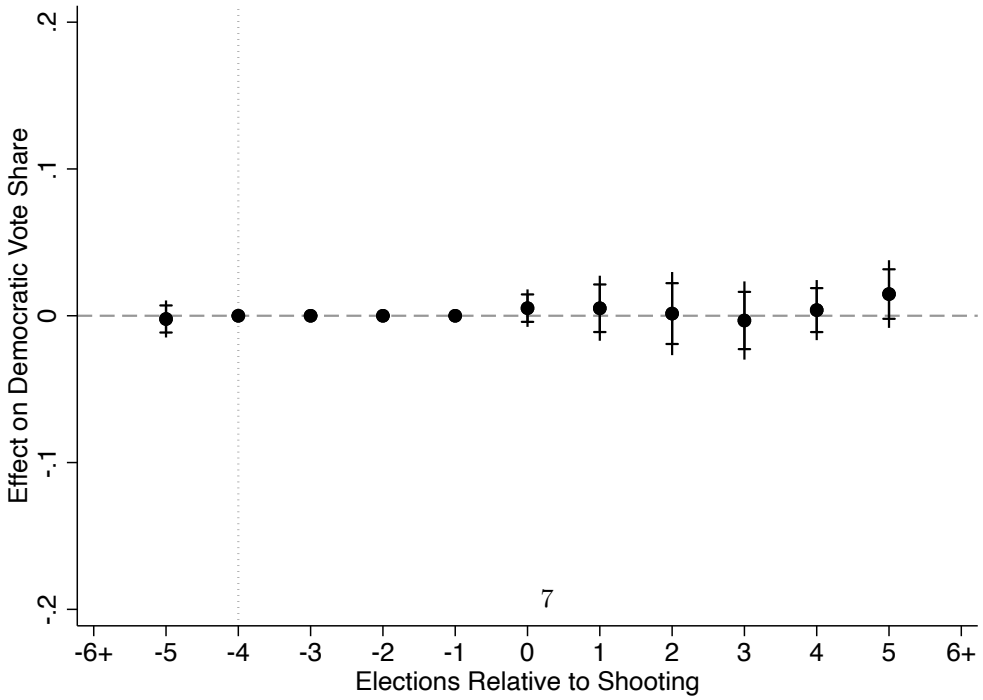
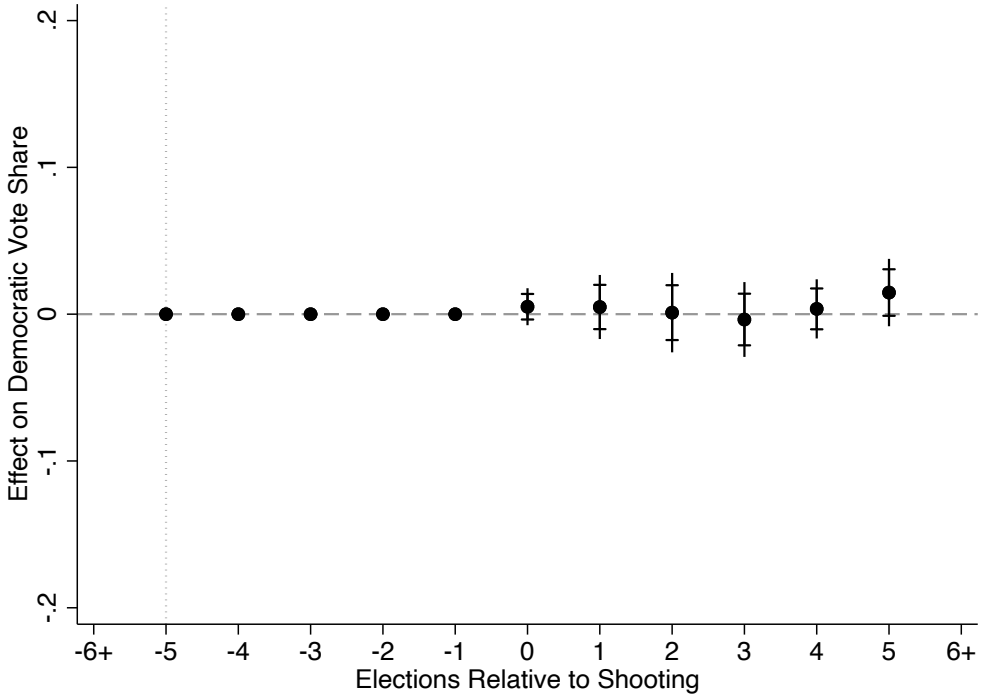





Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021).**Takeaway:** Benchmarked to pre-treatment trends at t-4, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

Figure S4: Results Alternate Baseline Periods in Event Study Design that Accounts for County Specific Time Trends (cont'd)

(a) LINEAR -5 Period Benchmark




(b) QUADRATIC -5 Period Benchmark

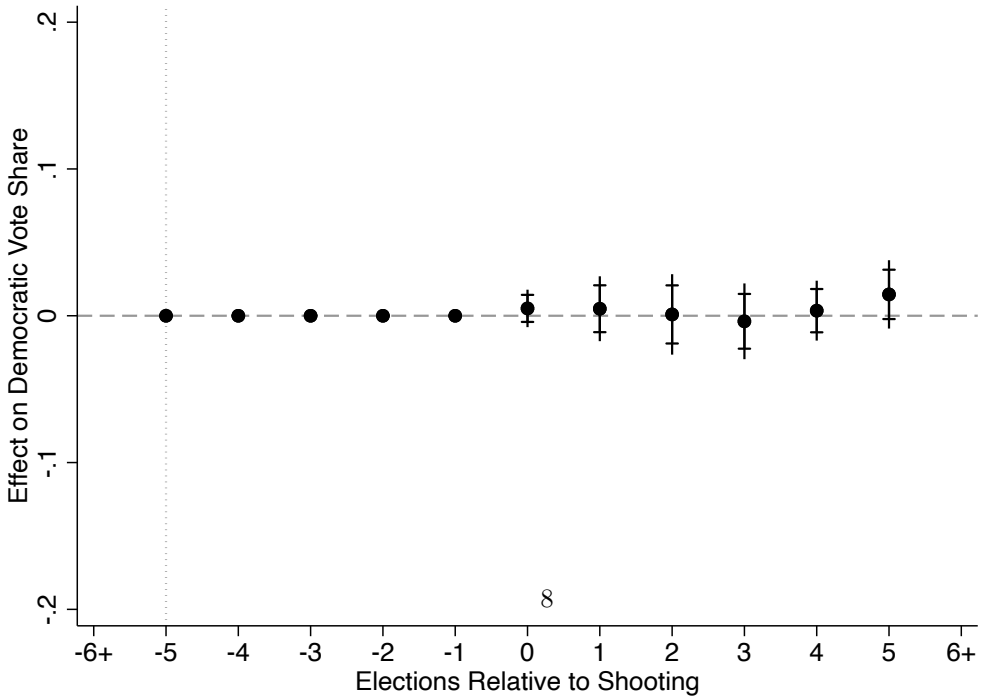
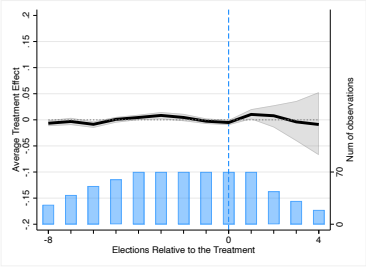




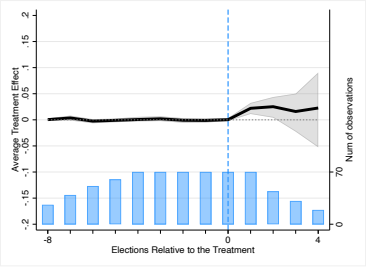

Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021).**Takeaway:** Benchmarked to pre-treatment trends at t-5, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

Figure S5: Interactive Fixed Effects Counterfactual Estimator

(a) Interactive Fixed Effects, r=2, degree=3




(b) Interactive Fixed Effects, r=3, degree=3




(c) Interactive Fixed Effects, r=1, degree=3

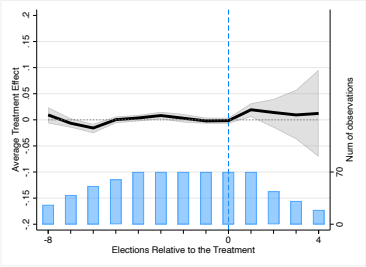
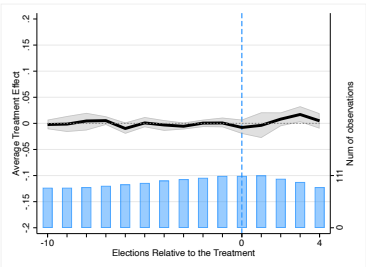


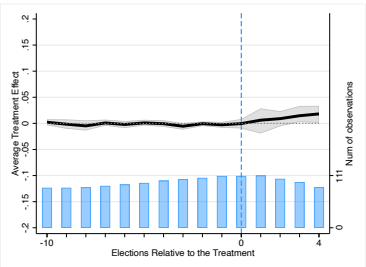

Figure shows the results from using Interactive Fixed Effects Counterfactual Estimator developed by Liu et al. (2021) with different values of $r$—the number of factors used in estimation—and the integer specifying the order of the polynomial trend term. **Takeaway:** In the interactive fixed effects models, there is no evidence of the substantial effects shown in more simplistic model specifications that do not account for potential violations of the parallel-trends assumption.

Figure S6: Interactive Fixed Effects Counterfactual Estimator (2)

(a) Interactive Fixed Effects, r=2, degree=2




(b) Interactive Fixed Effects, r=3, degree=2




(c) Interactive Fixed Effects, r=1, degree=2

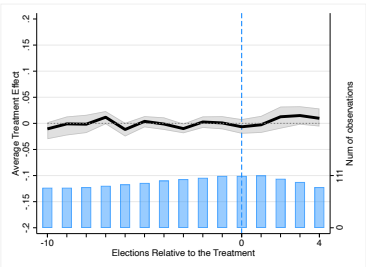
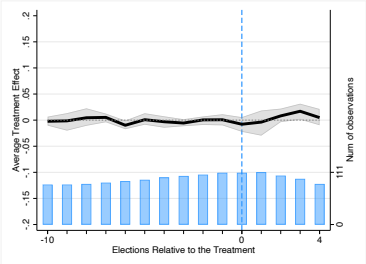


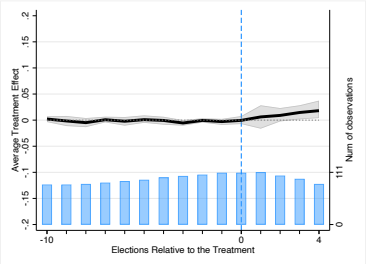

Figure shows the results from using Interactive Fixed Effects Counterfactual Estimator developed by Liu et al. (2021) with different values of $r$—the number of factors used in estimation—and the integer specifying the order of the polynomial trend term. **Takeaway:** In the interactive fixed effects models, there is no evidence of the substantial effects shown in more simplistic model specifications that do not account for potential violations of the parallel-trends assumption.

Figure S7: Interactive Fixed Effects Counterfactual Estimator (3)

(a) Interactive Fixed Effects, r=2, degree=4




(b) Interactive Fixed Effects, r=3, degree=4




(c) Interactive Fixed Effects, r=1, degree=4

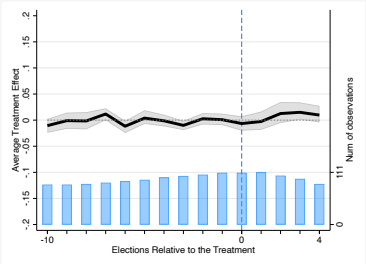
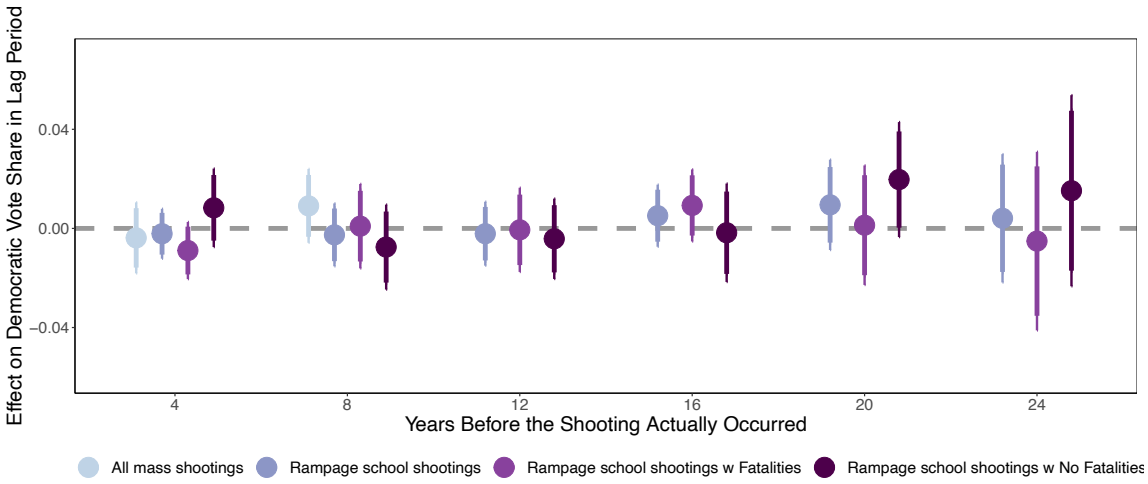



Figure shows the results from using Interactive Fixed Effects Counterfactual Estimator developed by Liu et al. (2021) with different values of $r$—the number of factors used in estimation—and the integer specifying the order of the polynomial trend term. **Takeaway:** In the interactive fixed effects models, there is no evidence of the substantial effects shown in more simplistic model specifications that do not account for potential violations of the parallel-trends assumption.

Figure S8: Pre-Treatment Effects with Alternate County-Specific Trend Types

(a) Cubic County Trends




(b) Quartic County Trends

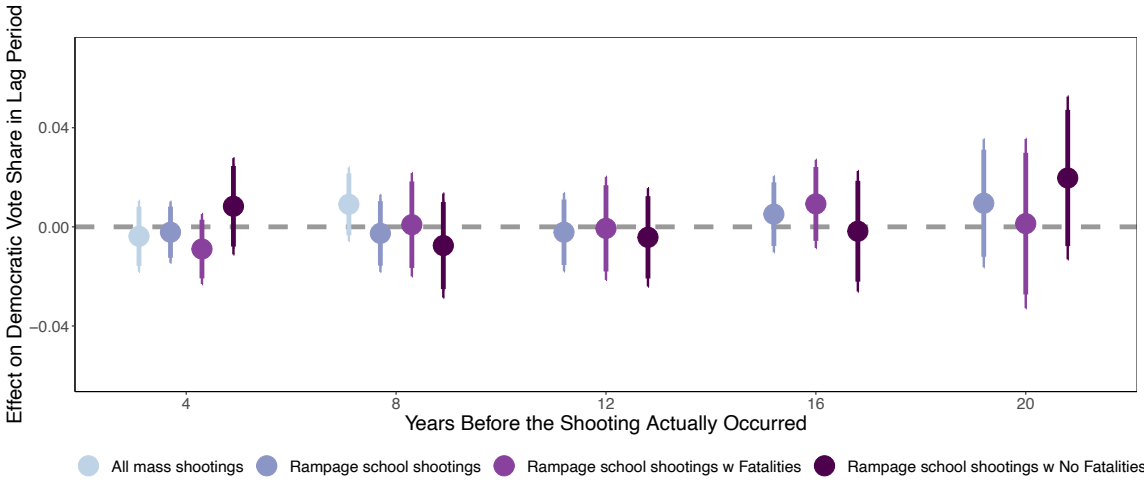
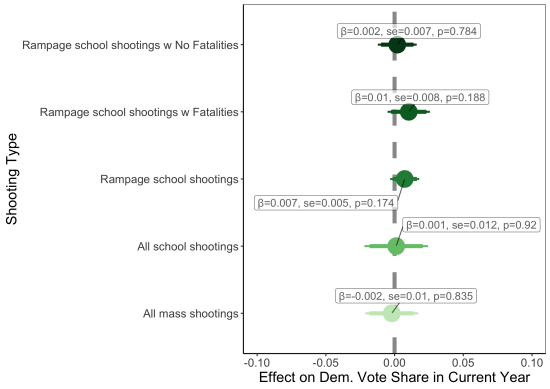


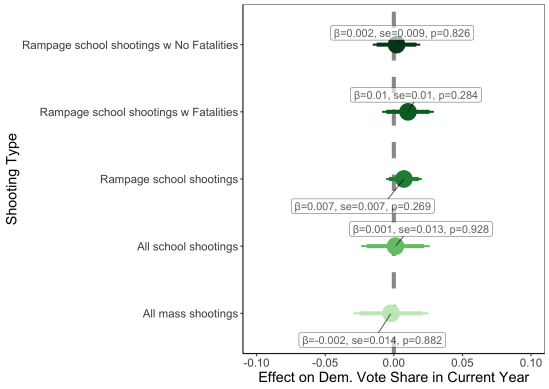

Figure shows the results from using higher order polynomial functional forms for the county-specific trends. The cubic trends model omits the 28 year lag because there are not enough observations in the GMAL data to estimate a model with this many high dimensional fixed effects. The quartic trends model omits the 24 and 28 year lag for the same reason. **Takeaway:** In contrast to the TWFE estimates shown in panels (a) and (b) in Figure 3 in the text (but consistent with specifications with linear and quadratic time trends), specifications with cubic and quadratic time trends show balance prior to when the shooting occurred.

Figure S9: The Effect of Mass Shootings on Presidential Election Returns Once County-Specific Trends are Absorbed, Alternate Polynomial Orders

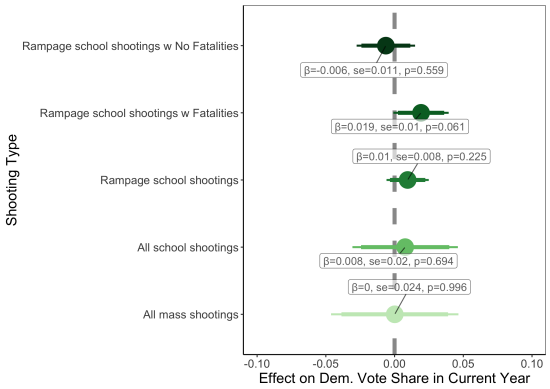
(a) Cubic County Trends Added




(b) Quartic County Trends Added




(c) Cubic County Trends Added, Change in DV




(d) Quartic County Trends Added, Change in DV

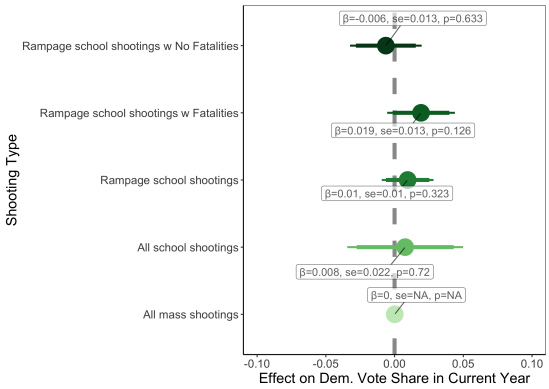
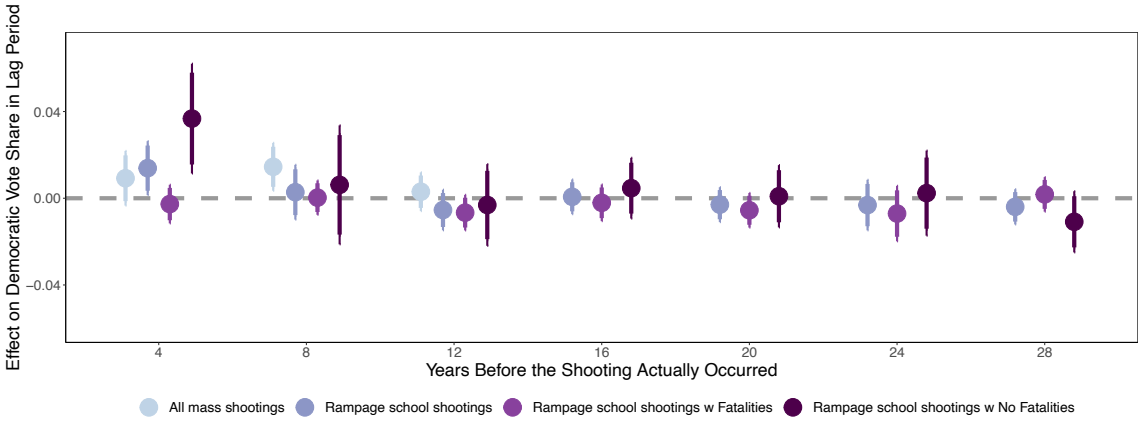


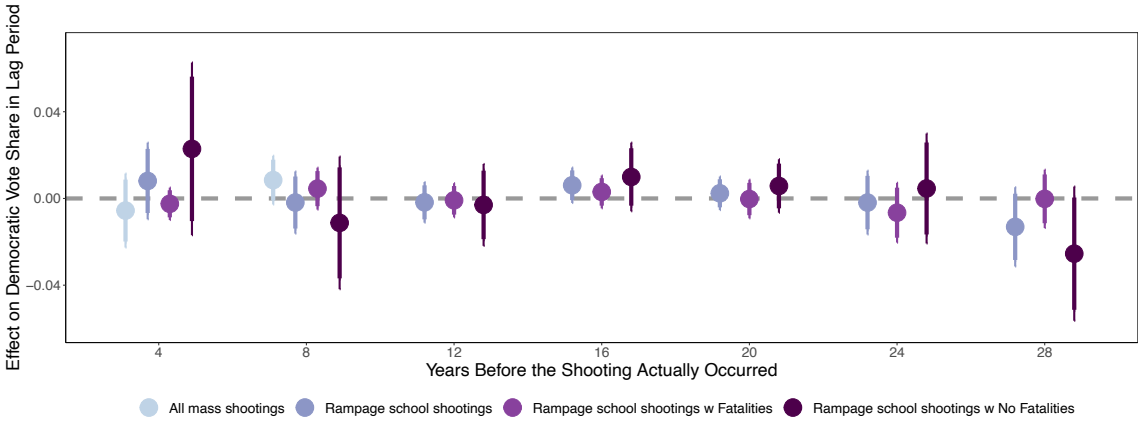

Figure shows the effect of mass shootings of various types once we account for differential trends in Democratic vote share across counties in the United States—this time with cubic and quartic county-specific trends. Within each panel, the first 3 estimates are using the GMAL coding of mass shootings and their data, the next comes from HHB, and the last comes from Yousaf. The upper left panel shows specifications with cubic county trends, the upper right panel shows specifications with quartic county trends, the bottom left panel shows specifications with cubic county trends and using a change in Democratic vote share over the prior 4-year-previous election, the bottom right panel shows specifications with quartic county trends and using a change in Democratic vote share over the prior 4-year-previous election. In the last panel, the standard error will not estimate for the Yousafdata as there are not observations in this shorter time series to do so. Coefficients, standard errors, and p-values are labeled for each coefficient. **Takeaway:** Once we account for differential trends across counties, the effects of mass shootings—be they located on school grounds or not, or be they rampage style or not—are all small and precisely-estimated.

Figure S10: Pre-Treatment Effects on Turnout

(a) TWFE




(b) Linear County Trends




(c) Quadratic County Trends

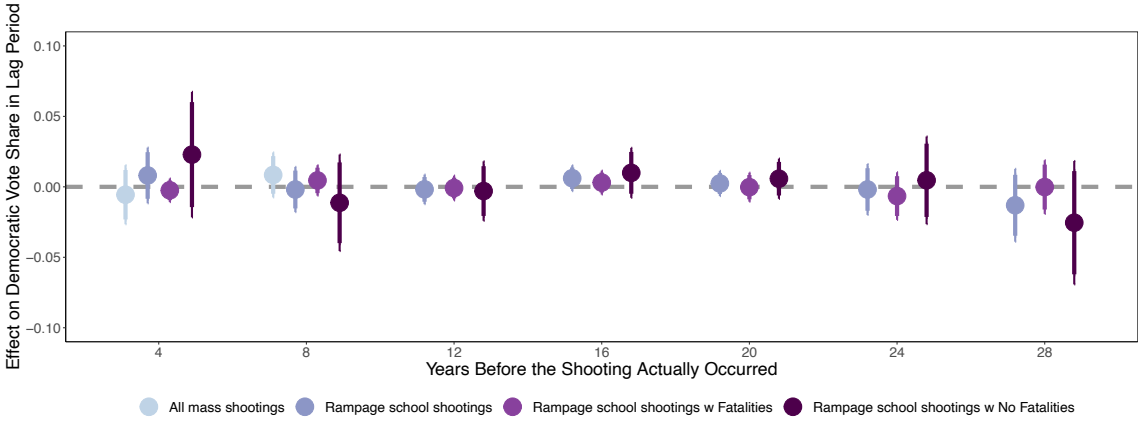
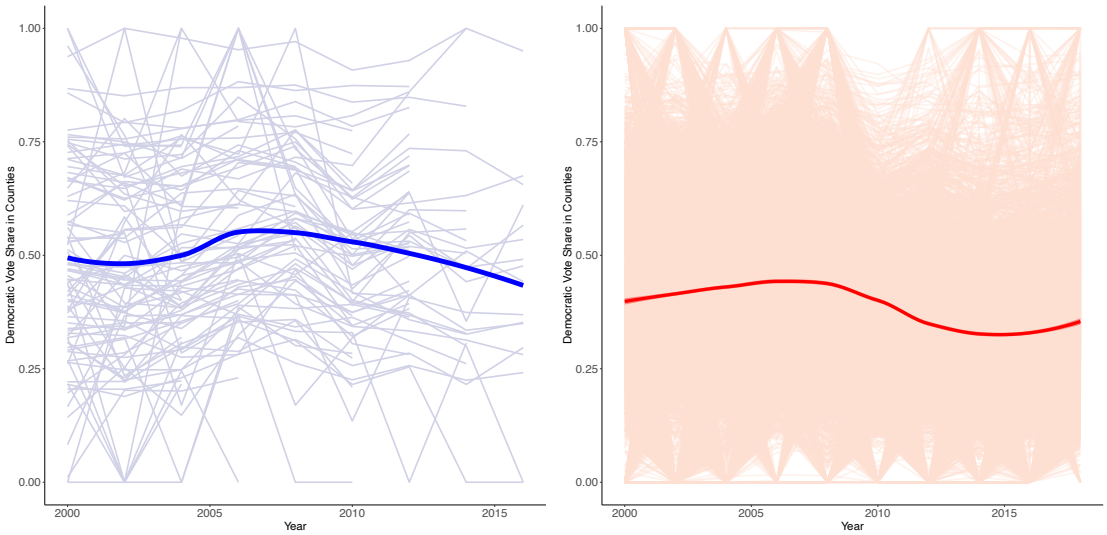



Figure shows the effect of mass shootings on voter turnout in the years prior to when a shooting occurred. **Takeaway:** In contrast to the effects of mass shootings on Democratic vote share which is plagued by trend differences pre-treatment, turnout does not appear to suffer from the same problem, as there is balance pre-treatment.

Figure S11: Trends in Presidential Vote Share in Counties With Mass Shootings Prior to These Shootings Occurring, Compared to Trends in Counties Without a Shooting (YOUSAF AND HHB DATA)

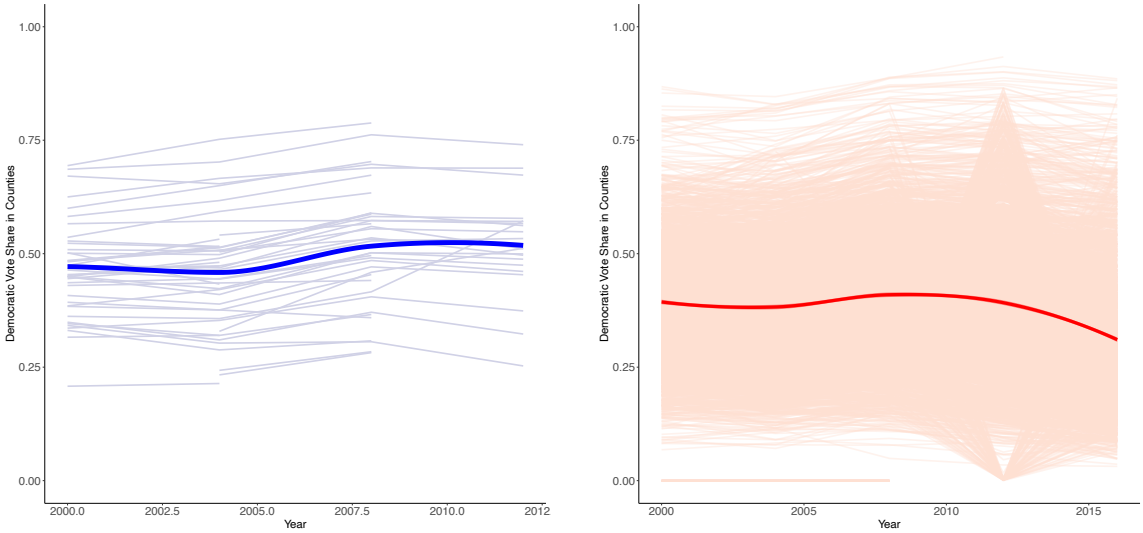
(a) Pre-treatment Trends in Democratic vote share in Shooting Counties (HHB)

(b) Trends in Democratic vote share in Non Shooting Counties (HHB)




(c) Pre-treatment Trends in Democratic vote share in Shooting Counties (Yousaf)

(d) Trends in Democratic vote share in Non Shooting Counties (Yousaf)



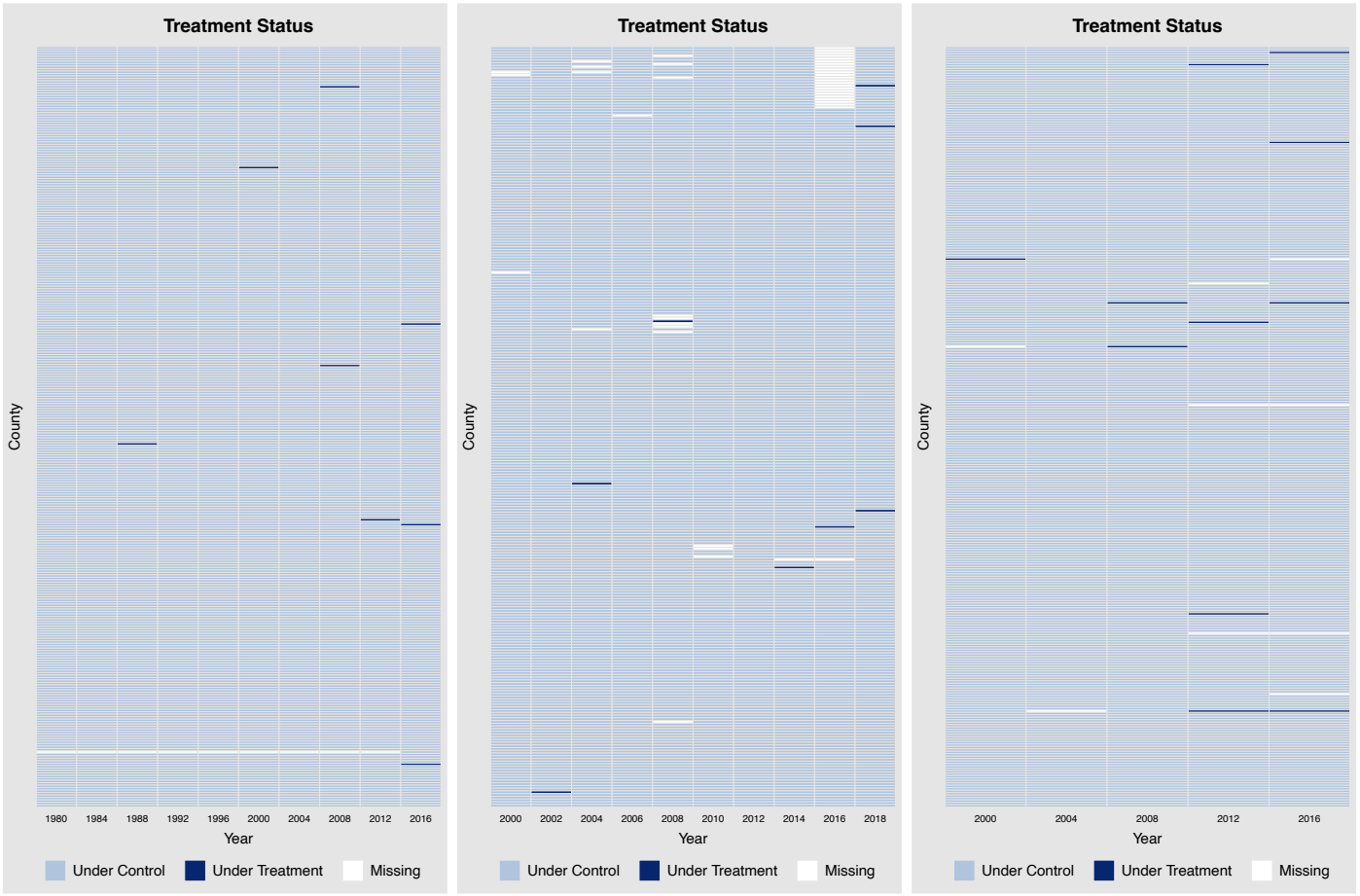

Pre-treatment trends of Democratic vote share in counties where a shooting occurred and benchmarks this to the trends in Democratic vote share found in counties where a shooting did not occur for the Yousaf and HHB data. In the panels on the left, the small blue lines mark the patterns for all counties with a shooting and the bolded **blue** lines capture the average trend across these counties. The panels on the right show the same pattern for counties without a shooting. The small red lines mark the patterns for all counties without a shooting and the bolded **red** lines shows a loess model for counties without a shooting. **Takeaway:** Though taking a slightly different shape that the GMAL data, both the HHB and YOUSAF datasets show a separation between pre-treatment counties and control counties.

Figure S12: Treatment Across Counties Over Time, Only County Years with a Shooting are Treated

(a) GMAL Treatment Panel for Random Sample

(b) HHB Treatment Panel for Random Sample

(c) Yousaf Treatment Panel for Random Sample



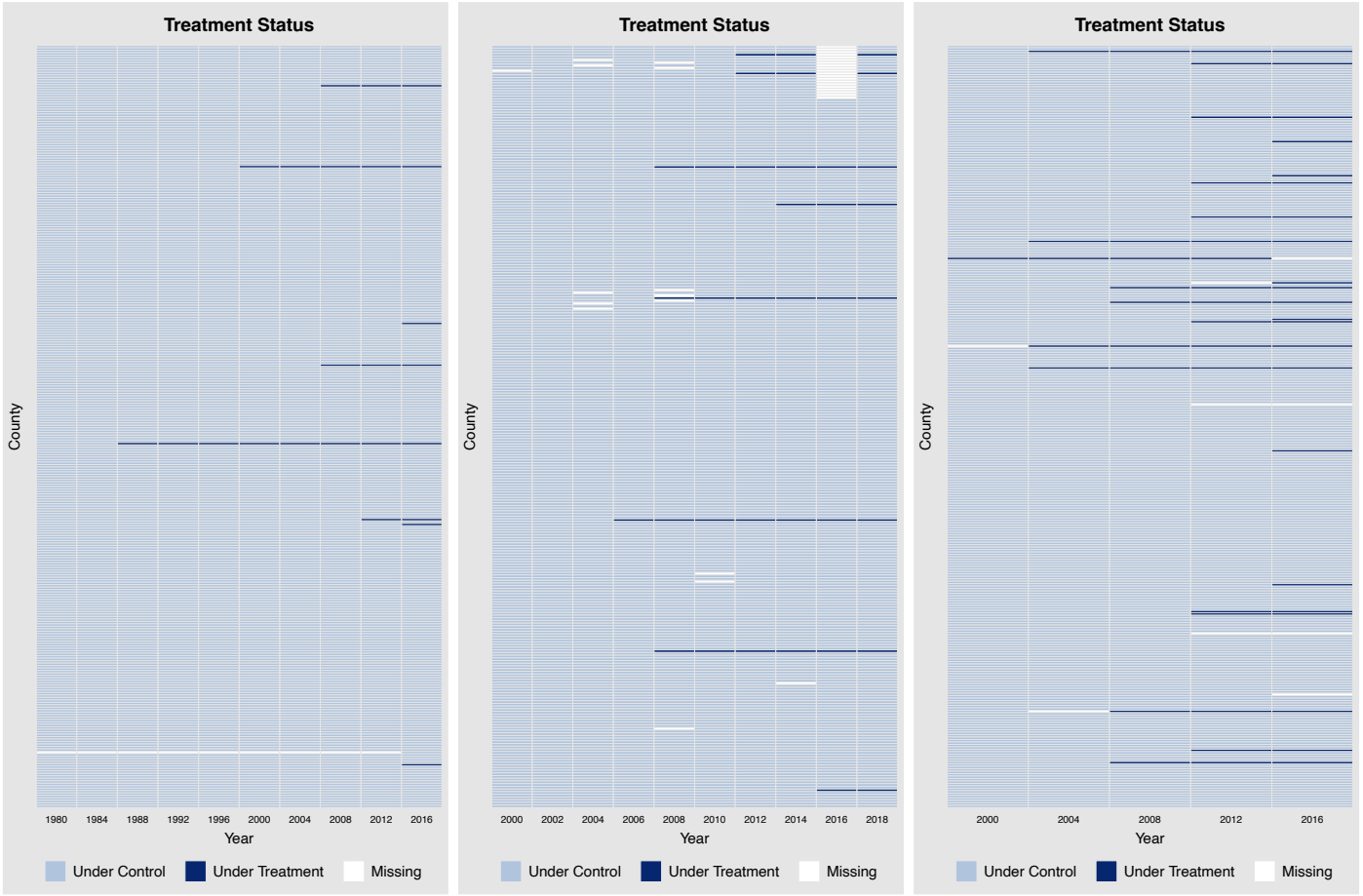

Treatment over time for a random sample of counties in the three datasets illustrating treatment approach 1. Separate random counties are used in the figure that follows this one.

Figure S13: Treatment Across Counties Over Time, All Post Shooting Counties are Treated

(a) GMAL Treatment Panel for Random Sample

(b) HHB Treatment Panel for Random Sample

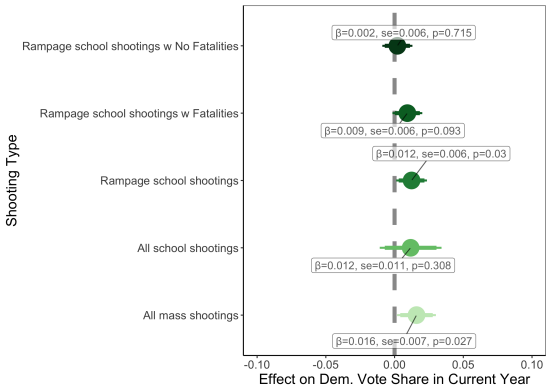
(c) Yousaf Treatment Panel for Random Sample



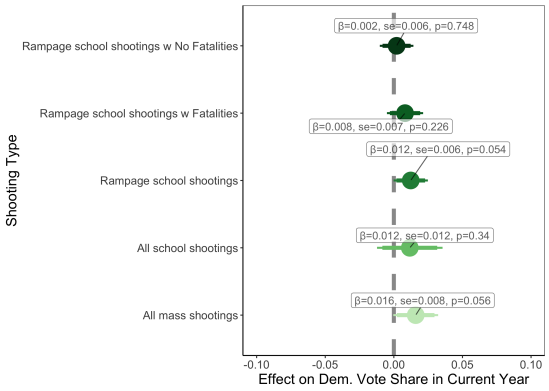

Treatment over time for a random sample of counties in the three datasets illustrating treatment approach 2. Separate random counties are used in the figure that precedes this one.

Figure S14: The Effect of Mass Shootings on Presidential Election Returns Once County-Specific Trends are Absorbed, All Post Shooting Counties are Treated

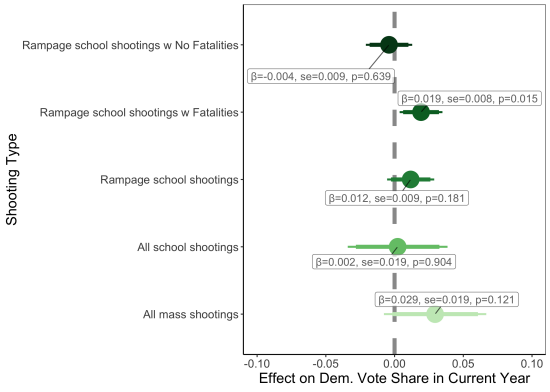
(a) Linear County Trends Added



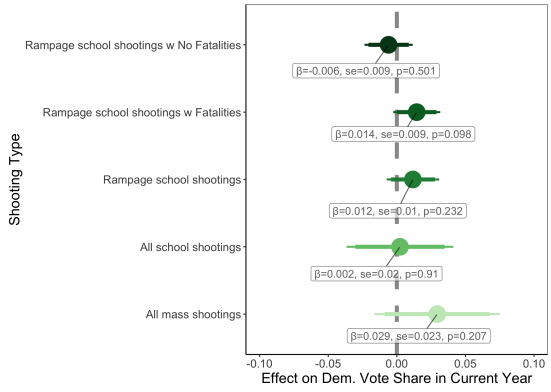

(b) Quadratic County Trends Added




(c) Linear County Trends Added, Change in DV




(d) Quadratic County Trends Added, Change in DV




sd

Effect of mass shootings of various types once we account for differential trends in Democratic vote share across counties in the United States. Within each panel, the first 3 estimates are using the GMAL coding of mass shootings and their data, the next comes from HHB, and the last comes from Yousaf. The upper left panel shows specifications with linear county trends, the upper right panel shows specifications with quadratic county trends, the bottom left panel shows specifications with linear county trends and using a change in Democratic vote share over the prior 4-year-previous election, the bottom right panel shows specifications with quadratic county trends and using a change in Democratic vote share over the prior 4-year-previous election. Coefficients, standard errors, and p-values are labeled for each coefficient. **Takeaway:** Once we account for differential trends across counties, the effects of mass shootings—be they located on school grounds or not, or be they rampage style or not—are all small and precisely-estimated.

Table S2: The ATT for each period, across all groups or cohorts (GMAL)

| stats | Average | T1984 | T1988 | T1992 | T1996 | T2000 | T2004 | T2008 | T2012 | T2016 |
|---|---|---|---|---|---|---|---|---|---|---|
| b | .0518245 | .0619944 | .0245811 | .0246506 | .0364387 | .053292 | .0565589 | .0617882 | .0599066 | .0872101 |
| se | .0096223 | .0136641 | .011968 | .01148 | .0120084 | .0125972 | .013094 | .0141372 | .0135578 | .0118054 |
| z | 5.385884 | 4.537023 | 2.053896 | 2.147266 | 3.034426 | 4.230466 | 4.319438 | 4.370606 | 4.418599 | 7.387303 |
| pvalue | 7.21e-08 | 5.71e-06 | .0399858 | .0317721 | .0024099 | .0000233 | .0000156 | .0000124 | 9.93e-06 | 1.50e-13 |
| ll | .0329652 | .0352133 | .0011242 | .0021502 | .0129026 | .028602 | .030895 | .0340798 | .0333337 | .064072 |
| ul | .0706839 | .0887756 | .0480381 | .0471509 | .0599748 | .0779821 | .0822227 | .0894967 | .0864794 | .1103483 |

Estimates of the ATT for each period, across all groups or cohorts (i.e. the "Calendar" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice.

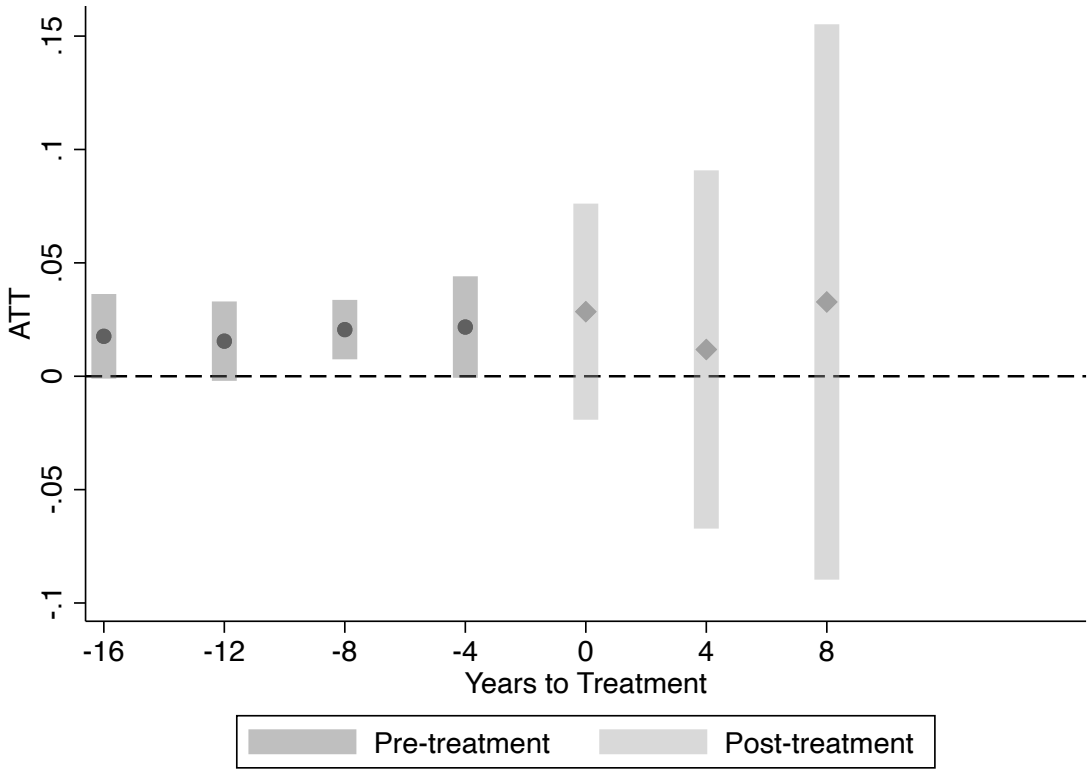


Table S3: The ATT for each group or cohort, across all periods (GMAL)

| stats | Average | G1984 | G1988 | G1992 | G1996 | G2000 | G2004 | G2008 | G2012 | G2016 |
|---|---|---|---|---|---|---|---|---|---|---|
| b | .053623 | .1587951 | .0391324 | .0485951 | .0563038 | .0283724 | .0285435 | .0474233 | .0469285 | .0565588 |
| se | .0068437 | .0189939 | .0201407 | .0162478 | .0424463 | .0300613 | .0270402 | .0116037 | .006478 | .0069854 |
| z | 7.835439 | 8.360334 | 1.942957 | 2.990872 | 1.326472 | .9438179 | 1.055595 | 4.086901 | 7.244247 | 8.096711 |
| pvalue | 4.67e-15 | 6.25e-17 | .0520213 | .0027818 | .1846836 | .3452627 | .2911533 | .0000437 | 4.35e-13 | 5.65e-16 |
| ll | .0402097 | .1215678 | -.0003425 | .01675 | -.0268894 | -.0305467 | -.0244543 | .0246804 | .0342318 | .0428676 |
| ul | .0670363 | .1960224 | .0786074 | .0804402 | .139497 | .0872915 | .0815413 | .0701663 | .0596252 | .0702499 |

Estimates of the ATT for each group or cohort, across all periods (i.e. the "Group" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice.

Figure S15: Estimation of all Dynamic Effects (GMAL)




Estimates of the dynamic effects (i.e. the "Event" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** Pre-treatment imbalances can be seen in the figure. This suggests that *even when* one uses "clean comparions" as suggested by Callaway and Sant'Anna (2021), differential pre-treatment trends are an issue. At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice. We reference the reader to the event study estimates in the paper for those that adjust for differential trends identified in the paper

Table S4: The ATT for each period, across all groups or cohorts (HHB)

| stats | Average | T2002 | T2004 | T2006 | T2008 | T2010 | T2012 | T2014 | T2016 | T2018 |
|---|---|---|---|---|---|---|---|---|---|---|
| b | .032836 | .051527 | .0243792 | -.0458838 | .0395123 | .0239998 | .0489557 | .0240271 | .0796322 | .0493749 |
| se | .0229896 | .0362721 | .0228559 | .0475194 | .0432784 | .0368888 | .0285933 | .0246188 | .0207751 | .0173127 |
| z | 1.4283 | 1.420568 | 1.066648 | -.9655811 | .9129784 | .650598 | 1.712138 | .9759645 | 3.833059 | 2.851938 |
| pvalue | .1532057 | .1554425 | .2861309 | .3342539 | .3612539 | .515306 | .0868712 | .3290821 | .0001266 | .0043454 |
| ll | -.0122227 | -.0195651 | -.0204175 | -.1390202 | -.0453119 | -.0483009 | -.0070862 | -.0242249 | .0389137 | .0154425 |
| ul | .0778948 | .1226191 | .0691759 | .0472525 | .1243364 | .0963005 | .1049976 | .072279 | .1203506 | .0833072 |

Estimates of the ATT for each period, across all groups or cohorts (i.e. the "Calendar" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice.

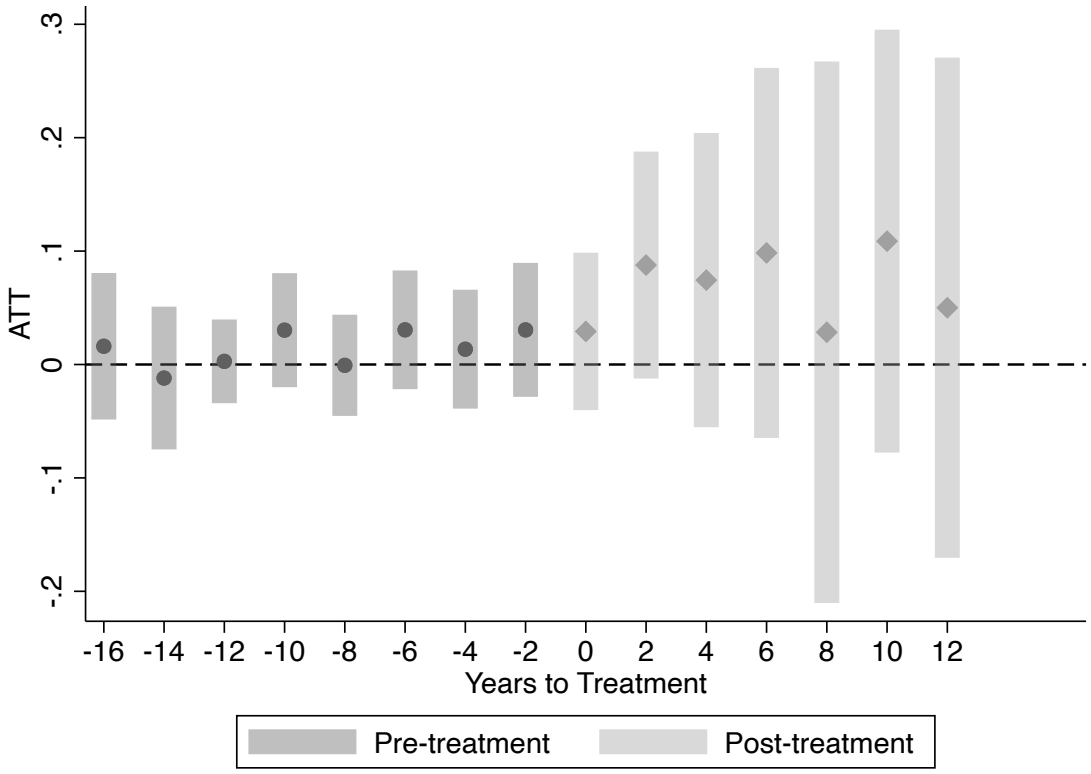


Table S5: The ATT for each group or cohort, across all periods (HHB)

| states | Average | G2002 | G2004 | G2006 | G2008 | G2010 | G2012 | G2014 | G2016 | G2018 |
|---|---|---|---|---|---|---|---|---|---|---|
| b | .0323525 | .0222296 | .056052 | -.0032929 | .1754662 | .137212 | .0905926 | .010359 | .000745 | .0107809 |
| se | .0142005 | .0621462 | .0256692 | .0714713 | .0603517 | .0051282 | .0259543 | .0148958 | .0178077 | .0251306 |
| z | 2.278258 | .3576986 | 2.183627 | -.0460732 | 2.907397 | 26.75635 | 3.490473 | .6954327 | .0418356 | .4289939 |
| pvalue | .0227112 | .7205689 | .0289897 | .9632519 | .0036445 | 1.0e-157 | .0004822 | .4867841 | .9666297 | .6679277 |
| ll | .0045199 | -.0995748 | .0057412 | -.1433741 | .0571791 | .1271609 | .0397232 | -.0188362 | -.0341574 | -.0384742 |
| ul | .060185 | .144034 | .1063627 | .1367882 | .2937533 | .1472631 | .141462 | .0395542 | .0356474 | .0600359 |

Estimates of the ATT for each group or cohort, across all periods (i.e. the "Group" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice.

Figure S16: Estimation of all Dynamic Effects (HHB)




Estimates of the dynamic effects (i.e. the "Event" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** Pretreatment imblances are of least concern in the HHB data, and this is where we observe no evidence for a significant effect. At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice. We reference the reader to the event study estimates in the paper for those that adjust for differential trends identified in the paper

Table S6: The ATT for each period, across all groups or cohorts (Yousaf)

| stats | Average | T2004 | T2008 | T2012 | T2016 |
|---|---|---|---|---|---|
| b | .0387695 | .015784 | .0336756 | .035597 | .0700214 |
| se | .0088303 | .0117028 | .0116784 | .0173425 | .0095523 |
| z | 4.390519 | 1.348739 | 2.88359 | 2.052591 | 7.330298 |
| pvalue | .0000113 | .1774209 | .0039317 | .0401123 | 2.30e-13 |
| ll | .0214625 | -.007153 | .0107865 | .0016064 | .0512992 |
| ul | .0560765 | .0387209 | .0565648 | .0695876 | .0887436 |

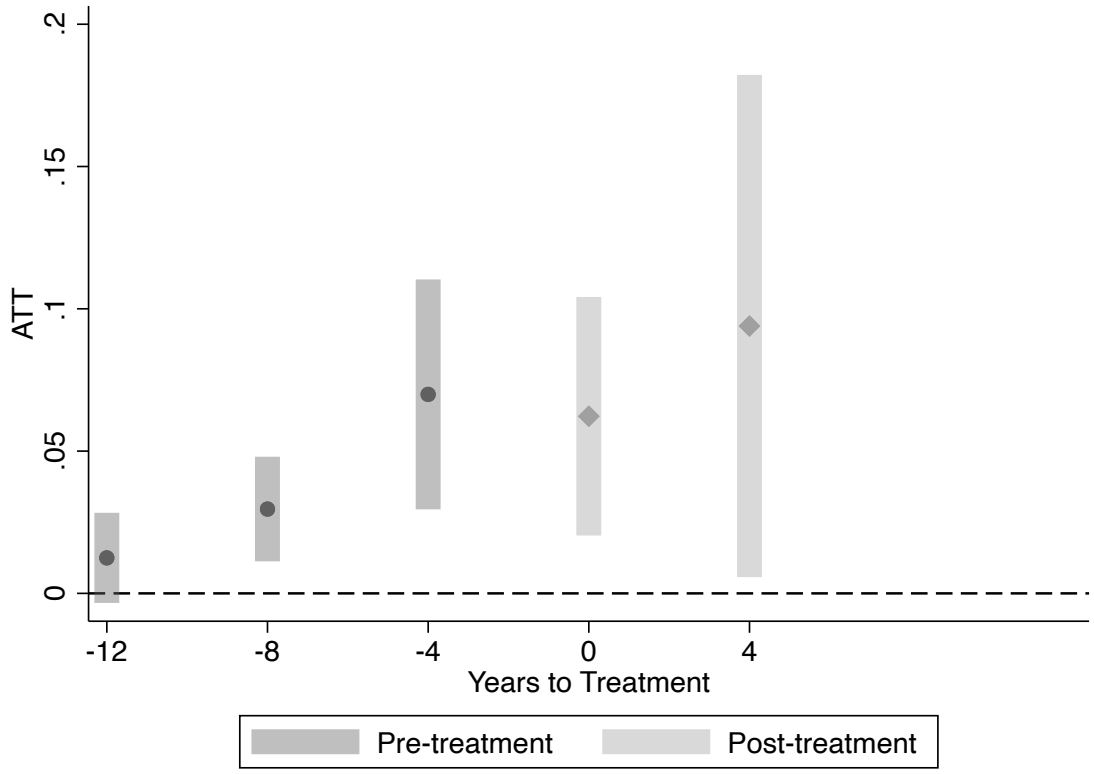
Estimates of the ATT for each period, across all groups or cohorts (i.e. the "Calendar" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice.

Table S7: The ATT for each group or cohort, across all periods (Yousaf)

| stats | Average | G2004 | G2008 | G2012 | G2016 |
|---|---|---|---|---|---|
| b | .0496422 | .0476684 | .0615136 | .0460837 | .0477795 |
| se | .0071074 | .0205126 | .0151624 | .0098212 | .0115534 |
| z | 6.984585 | 2.323858 | 4.056979 | 4.692281 | 4.135519 |
| pvalue | 2.86e-12 | .0201331 | .0000497 | 2.70e-06 | .0000354 |
| ll | .035712 | .0074644 | .0317958 | .0268345 | .0251351 |
| ul | .0635725 | .0878725 | .0912314 | .0653328 | .0704238 |

Estimates of the ATT for each group or cohort, across all periods (i.e. the "Group" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice.

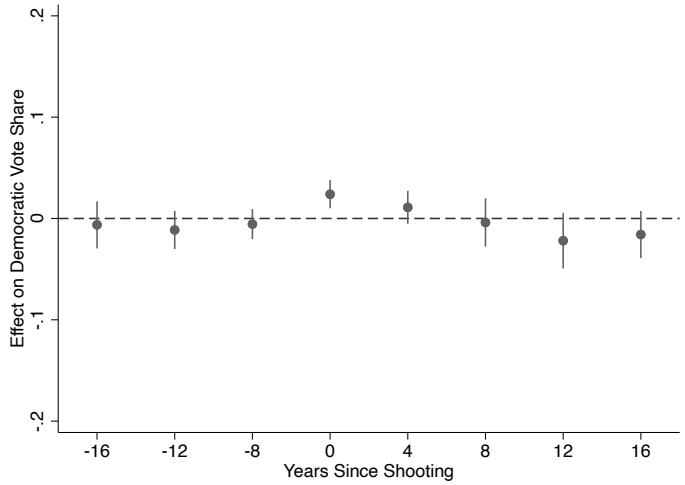
Figure S17: Estimation of all Dynamic Effects (Yousaf)



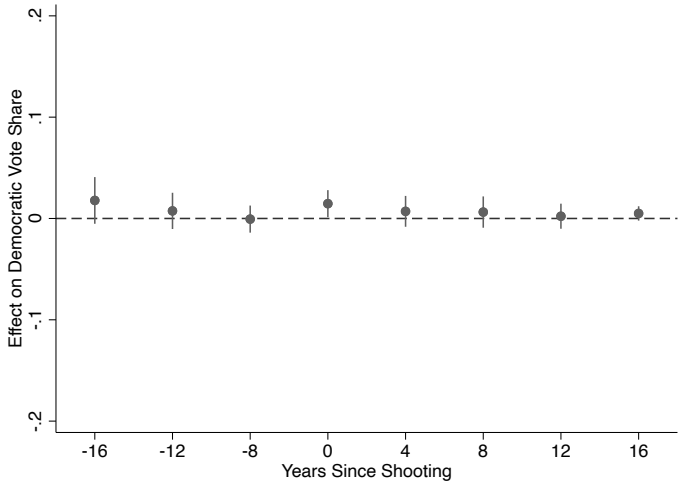

Estimates of the dynamic effects (i.e. the "Event" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** Pre-treatment imbalances can be seen in the figure. This suggests that *even when* one uses "clean comparions" as suggested by Callaway and Sant'Anna (2021), differential pre-treatment trends are an issue. At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice. We reference the reader to the event study estimates in the paper for those that adjust for differential trends identified in the paper

Figure S18: Sun and Abraham (2020) Event Study Estimates (GMAL)

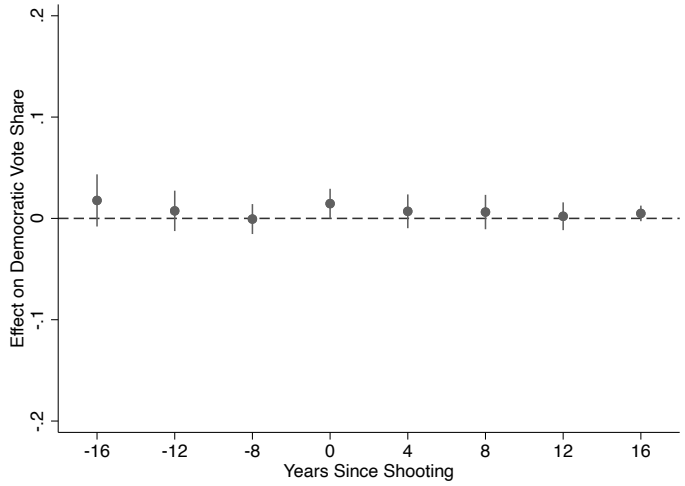
(a) TWFE




(b) Linear Trends



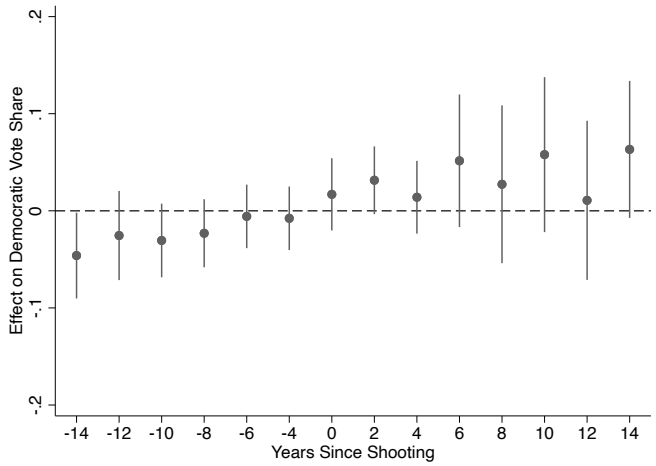

(c) Quadratic Trends



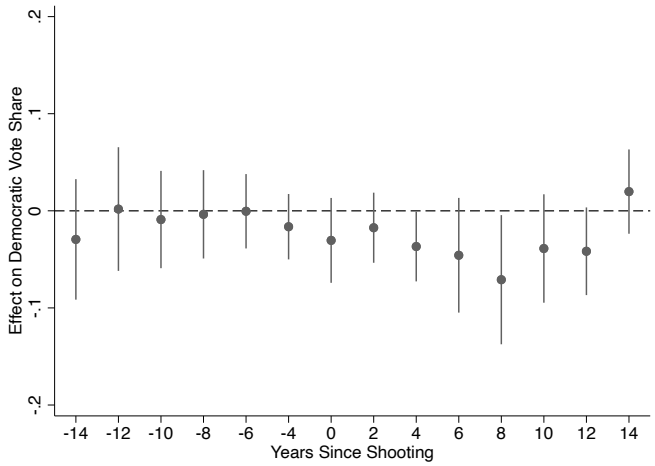

Sun and Abraham (2020) event study estimates through the `eventstudyinteract` package provided by the authors. Standard errors are clustered at the county level. **Takeaway:** Clean comparison effects with trends show no sign of a sizable and durable effect on Democratic vote shares shown in the TWFE nor in the simple event study plot

Figure S19: Sun and Abraham (2020) Event Study Estimates (HHB)

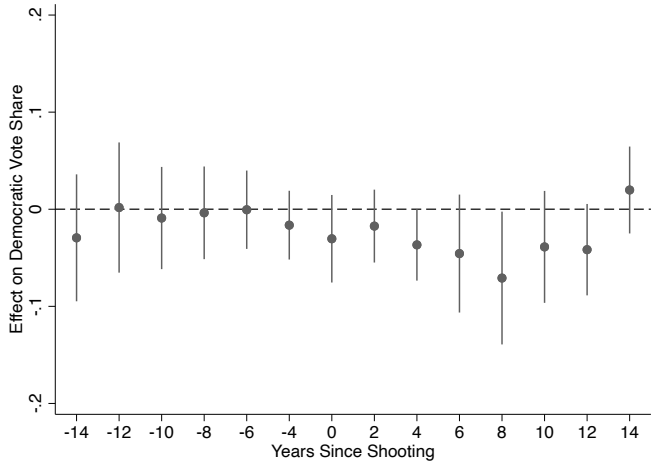
(a) TWFE




(b) Linear Trends



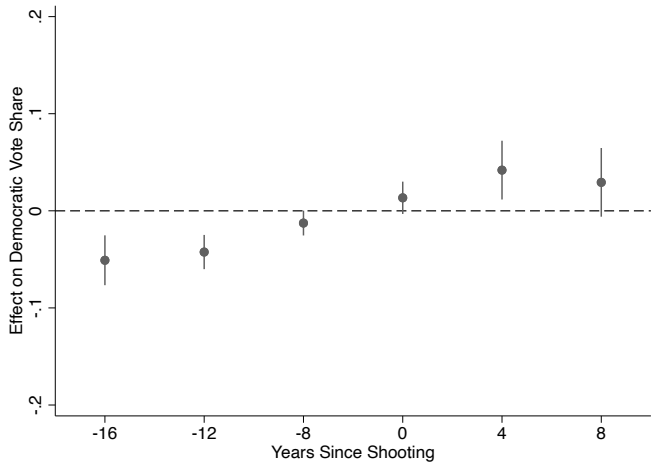

(c) Quadratic Trends



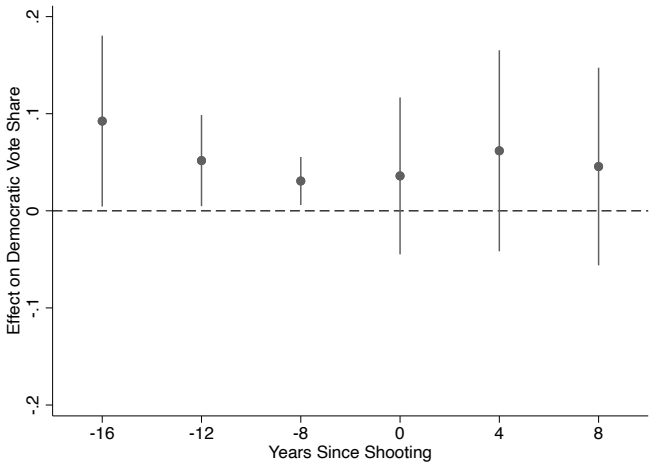

Sun and Abraham (2020) event study estimates through the `eventstudyinteract` package provided by the authors. Standard errors are clustered at the county level. **Takeaway:** Clean comparison effects with trends show no sign of a sizable and durable effect on Democratic vote shares shown in the TWFE nor in the simple event study plot

Figure S20: Sun and Abraham (2020) Event Study Estimates (Yousaf)

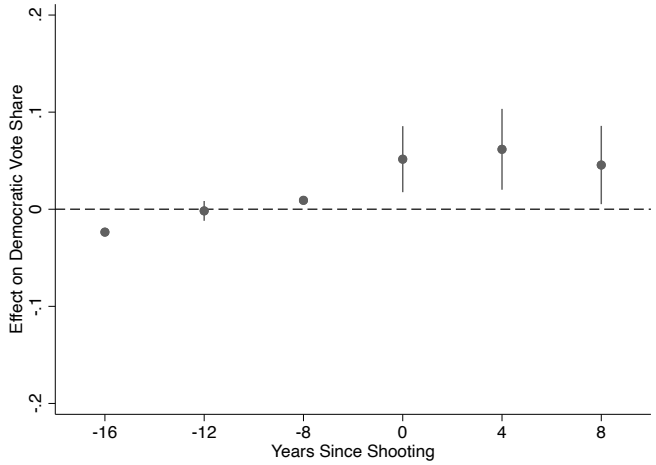
(a) TWFE




(b) Linear Trends

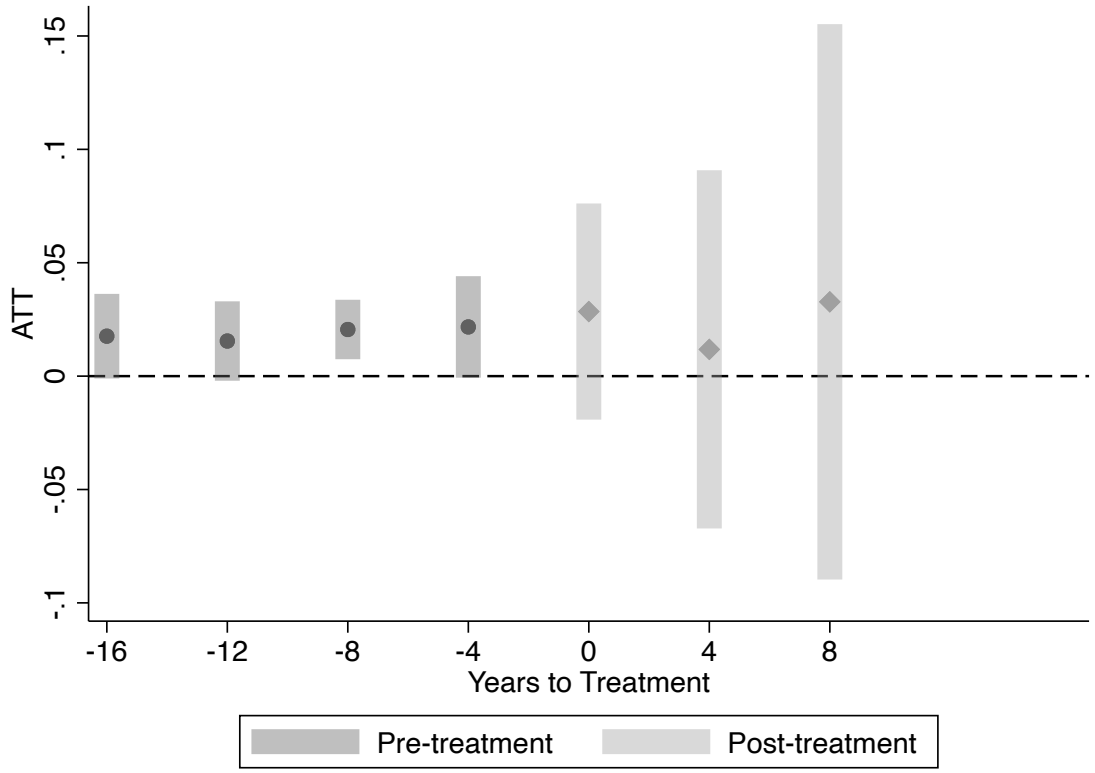



(c) Quadratic Trends




Sun and Abraham (2020) event study estimates through the `eventstudyinteract` package provided by the authors. Standard errors are clustered at the county level. **Takeaway:** Clean comparison effects with trends show no robust sign of a sizable and durable effect on Democratic vote shares shown in the TWFE nor in the simple event study plot. The trends specification for this approach in the Yousaf data still show signs of pre-treatment imbalance and, as such, should be interpreted with care.

Figure S21: Estimation of Clean Comparison TWFE Effects using the Callaway and Sant'Anna (2021) Approach




Estimates of the dynamic effects (i.e. the "Event" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. Controls used by GMAL are included—i.e. population, proportion non-white, and change in the unemployment rate. **Takeaway:** Pre-treatment imbalances can still be seen in the figure. This suggests that *even when* one uses "clean comparions" as suggested by Callaway and Sant'Anna (2021) and covariates, differential pre-treatment trends are an issue. At present, this method does not allow for the inclusion of unit-present trends, so estimates for this empirical case should be viewed with caution. These are included as an illustration of how to use this method in practice. We reference the reader to the event study estimates in the paper for those that adjust for differential trends identified in the paper.

Table S8: Estimation of Clean Comparison TWFE Effects using the de Chaisemartin and D'Haultfoeuille Approach – HHB

| | Estimate | SE | LB CI | UB CI | N | Switchers |
|---|---|---|---|---|---|---|
| Effect_0 | 0.0124381 | 0.0186208 | -0.0240587 | 0.048935 | 26791 | 91 |
| Effect_1 | 0.0180224 | 0.0268722 | -0.0346472 | 0.0706919 | 23526 | 71 |
| Effect_2 | 0.0051758 | 0.0350417 | -0.063506 | 0.0738576 | 20319 | 61 |
| Effect_3 | 0.0493664 | 0.0762519 | -0.1000873 | 0.1988201 | 17146 | 34 |
| Effect_4 | 0.0116177 | 0.1171154 | -0.2179286 | 0.2411639 | 14088 | 26 |
| Placebo_1 | 0.016137 | 0.0113333 | -0.0060763 | 0.0383503 | 23540 | 85 |
| Placebo_2 | -0.0214889 | 0.0152387 | -0.0513567 | 0.0083788 | 20341 | 83 |
| Placebo_3 | 0.0211926 | 0.0126292 | -0.0035607 | 0.0459458 | 17184 | 72 |
| Placebo_4 | -0.0103004 | 0.0117131 | -0.0332582 | 0.0126574 | 14126 | 64 |

*Note:* de Chaisemartin and D'Haultfoeuille (2020) approach for assessing and addressing implemented in the `did_multiplegt` package in *STATA* and `DIDmultiplegt` package in *R*. Under the common trends assumption, beta estimates a weighted sum of 395 ATTs. 379 ATTs receive a positive weight, and 16 receive a negative weight. The sum of the positive weights is equal to 1.0010116. The sum of the negative weights is equal to -.00101162. beta is compatible with a DGP where the average of those ATTs is equal to 0, while their standard deviation is equal to .12133344. beta is compatible with a DGP where those ATTs all are of a different sign than beta, while their standard deviation is equal to 13.249181. **Takeaway:** After using this method, we see no evidence of substantial or significant effects of mass shootings on electoral outcomes.

Table S9: Estimation of Clean Comparison TWFE Effects using the de Chaisemartin and D'Haultfoeuille Approach – GMAL

| | Estimate | SE | LB CI | UB CI | N | Switchers |
|---|---|---|---|---|---|---|
| Effect_0 | 0.0170459 | 0.004788 | 0.0076614 | 0.0264304 | 27632 | 98 |
| Effect_1 | 0.0159759 | 0.0101364 | -0.0038914 | 0.0358431 | 24507 | 72 |
| Effect_2 | 0.0227205 | 0.0188645 | -0.0142539 | 0.0596949 | 21409 | 59 |
| Effect_3 | 0.0359566 | 0.0292597 | -0.0213925 | 0.0933056 | 18315 | 47 |
| Effect_4 | 0.0797536 | 0.0415763 | -0.0017359 | 0.1612432 | 15231 | 38 |
| Placebo_1 | 0.0000278 | 0.004569 | -0.0089274 | 0.0089831 | 24526 | 91 |
| Placebo_2 | 0.006823 | 0.0051979 | -0.0033649 | 0.0170109 | 21429 | 79 |
| Placebo_3 | -0.0031506 | 0.0038482 | -0.010693 | 0.0043917 | 18344 | 76 |

*Note:* de Chaisemartin and D'Haultfoeuille (2020) approach for assessing and addressing implemented in the `did_multiplegt` package in *STATA* and `DIDmultiplegt` package in *R*. Under the common trends assumption, beta estimates a weighted sum of 400 ATTs. 396 ATTs receive a positive weight, and 4 receive a negative weight. The sum of the positive weights is equal to 1.0002424. The sum of the negative weights is equal to -.00024243. beta is compatible with a DGP where the average of those ATTs is equal to 0, while their standard deviation is equal to .13523432. beta is compatible with a DGP where those ATTs all are of a different sign than beta, while their standard deviation is equal to 30.161087. **Takeaway:** After using this method, we see no evidence of substantial or significant effects of mass shootings on electoral outcomes. Effect_0 is not robust to other approaches for adjusting for potential violations of the parallel trends assumption—e.g. Rambachan and Roth (2021).

# Figure Tables

Table S10: Figure 1a Results

| treatment | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|
| YOUSAF | .029 | .008 | 3.634 | 0 | 15570 | .81 |
| GMAL | .055 | .009 | 5.964 | 0 | 31040 | .762 |
| HHB | .026 | .011 | 2.493 | .013 | 30625 | .657 |

Table S11: Figure 1b Results

| treatment | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|
| YOUSAF | .058 | .005 | 12.382 | 0 | 15570 | .812 |
| GMAL | .087 | .012 | 7.493 | 0 | 31040 | .763 |
| HHB | .078 | .015 | 5.303 | 0 | 30625 | .657 |

Table S12: Figure 3a Results

| treatment | study | Years Prior | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|---|---|
| All mass shootings | YOUSAF | 12 | .002 | .006 | .413 | .679 | 6132 | .971 |
| All mass shootings | YOUSAF | 8 | .016 | .006 | 2.723 | .006 | 9270 | .932 |
| All mass shootings | YOUSAF | 4 | .014 | .007 | 2.088 | .037 | 12432 | .796 |
| Rampage school No Fatalities | GMAL | 28 | -.012 | .012 | -.974 | .33 | 9305 | .84 |
| Rampage school No Fatalities | GMAL | 24 | .017 | .009 | 1.842 | .066 | 12408 | .849 |
| Rampage school No Fatalities | GMAL | 20 | .03 | .007 | 4.493 | 0 | 15511 | .854 |
| Rampage school No Fatalities | GMAL | 16 | .025 | .01 | 2.381 | .017 | 18614 | .843 |
| Rampage school No Fatalities | GMAL | 12 | .026 | .01 | 2.511 | .012 | 18616 | .881 |
| Rampage school No Fatalities | GMAL | 8 | .032 | .01 | 3.297 | .001 | 18618 | .841 |
| Rampage school No Fatalities | GMAL | 4 | .041 | .012 | 3.487 | 0 | 18619 | .856 |
| Rampage school Fatalities | GMAL | 28 | .007 | .01 | .673 | .501 | 9305 | .84 |
| Rampage school Fatalities | GMAL | 24 | .005 | .007 | .77 | .442 | 12408 | .849 |
| Rampage school Fatalities | GMAL | 20 | .015 | .008 | 1.866 | .062 | 15511 | .854 |
| Rampage school Fatalities | GMAL | 16 | .025 | .008 | 3.185 | .001 | 18614 | .843 |
| Rampage school Fatalities | GMAL | 12 | .021 | .009 | 2.402 | .016 | 18616 | .881 |
| Rampage school Fatalities | GMAL | 8 | .033 | .01 | 3.309 | .001 | 18618 | .841 |
| Rampage school Fatalities | GMAL | 4 | .027 | .009 | 2.806 | .005 | 18619 | .856 |
| Rampage school shootings | GMAL | 28 | -.001 | .008 | -.173 | .863 | 9305 | .84 |
| Rampage school shootings | GMAL | 24 | .011 | .006 | 1.877 | .061 | 12408 | .849 |
| Rampage school shootings | GMAL | 20 | .022 | .005 | 4.075 | 0 | 15511 | .854 |
| Rampage school shootings | GMAL | 16 | .026 | .006 | 4.176 | 0 | 18614 | .844 |
| Rampage school shootings | GMAL | 12 | .024 | .007 | 3.507 | 0 | 18616 | .881 |
| Rampage school shootings | GMAL | 8 | .034 | .007 | 4.778 | 0 | 18618 | .841 |
| Rampage school shootings | GMAL | 4 | .034 | .007 | 4.534 | 0 | 18619 | .856 |

Table S13: Figure 3b Results

| treatment | study | Years Prior | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|---|---|
| All mass shootings | YOUSAF | 12 | .01 | .004 | 2.436 | .015 | 6132 | .971 |
| All mass shootings | YOUSAF | 8 | .02 | .005 | 4.078 | 0 | 9270 | .932 |
| All mass shootings | YOUSAF | 4 | .027 | .006 | 4.482 | 0 | 12432 | .797 |
| Rampage school No Fatalities | GMAL | 28 | -.012 | .012 | -.974 | .33 | 9305 | .84 |
| Rampage school No Fatalities | GMAL | 24 | .017 | .009 | 1.842 | .066 | 12408 | .849 |
| Rampage school No Fatalities | GMAL | 20 | .03 | .007 | 4.493 | 0 | 15511 | .854 |
| Rampage school No Fatalities | GMAL | 16 | .025 | .01 | 2.381 | .017 | 18614 | .843 |
| Rampage school No Fatalities | GMAL | 12 | .026 | .01 | 2.511 | .012 | 18616 | .881 |
| Rampage school No Fatalities | GMAL | 8 | .032 | .01 | 3.297 | .001 | 18618 | .841 |
| Rampage school No Fatalities | GMAL | 4 | .041 | .012 | 3.487 | 0 | 18619 | .856 |
| Rampage school Fatalities | GMAL | 28 | .006 | .01 | .56 | .575 | 9305 | .84 |
| Rampage school Fatalities | GMAL | 24 | .003 | .007 | .441 | .659 | 12408 | .849 |
| Rampage school Fatalities | GMAL | 20 | .014 | .008 | 1.799 | .072 | 15511 | .854 |
| Rampage school Fatalities | GMAL | 16 | .025 | .008 | 3.249 | .001 | 18614 | .843 |
| Rampage school Fatalities | GMAL | 12 | .021 | .008 | 2.511 | .012 | 18616 | .881 |
| Rampage school Fatalities | GMAL | 8 | .033 | .01 | 3.39 | .001 | 18618 | .841 |
| Rampage school Fatalities | GMAL | 4 | .026 | .009 | 2.874 | .004 | 18619 | .856 |
| Rampage school shootings | GMAL | 28 | .013 | .01 | 1.3 | .194 | 9305 | .84 |
| Rampage school shootings | GMAL | 24 | .031 | .006 | 5.429 | 0 | 12408 | .849 |
| Rampage school shootings | GMAL | 20 | .039 | .006 | 6.817 | 0 | 15511 | .855 |
| Rampage school shootings | GMAL | 16 | .043 | .007 | 6.096 | 0 | 18614 | .844 |
| Rampage school shootings | GMAL | 12 | .048 | .007 | 7.324 | 0 | 18616 | .881 |
| Rampage school shootings | GMAL | 8 | .066 | .008 | 8.174 | 0 | 18618 | .842 |
| Rampage school shootings | GMAL | 4 | .063 | .01 | 6.376 | 0 | 18619 | .857 |

Table S14: Figure 3c Results

| treatment | study | Years Prior | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|---|---|
| All mass shootings | YOUSAF | 8 | .009 | .005 | 1.699 | .089 | 9270 | .991 |
| All mass shootings | YOUSAF | 4 | -.004 | .006 | -.64 | .522 | 12432 | .912 |
| Rampage school No Fatalities | GMAL | 28 | -.007 | .015 | -.433 | .665 | 9305 | .975 |
| Rampage school No Fatalities | GMAL | 24 | .015 | .011 | 1.343 | .18 | 12408 | .91 |
| Rampage school No Fatalities | GMAL | 20 | .02 | .008 | 2.357 | .019 | 15511 | .909 |
| Rampage school No Fatalities | GMAL | 16 | -.002 | .008 | -.215 | .83 | 18614 | .91 |
| Rampage school No Fatalities | GMAL | 12 | -.004 | .006 | -.638 | .523 | 18616 | .945 |
| Rampage school No Fatalities | GMAL | 8 | -.007 | .007 | -1.101 | .271 | 18618 | .942 |
| Rampage school No Fatalities | GMAL | 4 | .008 | .006 | 1.34 | .18 | 18619 | .973 |
| Rampage school Fatalities | GMAL | 28 | -.015 | .015 | -.999 | .318 | 9305 | .975 |
| Rampage school Fatalities | GMAL | 24 | -.005 | .011 | -.483 | .629 | 12408 | .91 |
| Rampage school Fatalities | GMAL | 20 | .001 | .009 | .15 | .881 | 15511 | .909 |
| Rampage school Fatalities | GMAL | 16 | .009 | .006 | 1.622 | .105 | 18614 | .91 |
| Rampage school Fatalities | GMAL | 12 | 0 | .007 | -.072 | .942 | 18616 | .945 |
| Rampage school Fatalities | GMAL | 8 | .001 | .007 | .148 | .883 | 18618 | .942 |
| Rampage school Fatalities | GMAL | 4 | -.009 | .005 | -1.955 | .051 | 18619 | .973 |
| Rampage school shootings | GMAL | 28 | -.011 | .011 | -1.031 | .303 | 9305 | .975 |
| Rampage school shootings | GMAL | 24 | .004 | .008 | .537 | .591 | 12408 | .91 |
| Rampage school shootings | GMAL | 20 | .01 | .007 | 1.449 | .147 | 15511 | .909 |
| Rampage school shootings | GMAL | 16 | .005 | .005 | 1.046 | .296 | 18614 | .91 |
| Rampage school shootings | GMAL | 12 | -.002 | .005 | -.398 | .691 | 18616 | .945 |
| Rampage school shootings | GMAL | 8 | -.003 | .005 | -.5 | .617 | 18618 | .942 |
| Rampage school shootings | GMAL | 4 | -.002 | .004 | -.529 | .597 | 18619 | .973 |

Table S15: Figure 3d Results

| treatment | study | Years Prior | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|---|---|
| All mass shootings | YOUSAF | 8 | .002 | .006 | .262 | .794 | 9270 | .991 |
| All mass shootings | YOUSAF | 4 | -.005 | .009 | -.553 | .58 | 12432 | .912 |
| Rampage school No Fatalities | GMAL | 28 | -.007 | .015 | -.488 | .625 | 9305 | .975 |
| Rampage school No Fatalities | GMAL | 24 | .016 | .011 | 1.422 | .155 | 12408 | .91 |
| Rampage school No Fatalities | GMAL | 20 | .02 | .008 | 2.444 | .015 | 15511 | .909 |
| Rampage school No Fatalities | GMAL | 16 | -.001 | .008 | -.089 | .929 | 18614 | .91 |
| Rampage school No Fatalities | GMAL | 12 | -.004 | .006 | -.628 | .53 | 18616 | .945 |
| Rampage school No Fatalities | GMAL | 8 | -.008 | .007 | -1.18 | .238 | 18618 | .942 |
| Rampage school No Fatalities | GMAL | 4 | .007 | .006 | 1.181 | .238 | 18619 | .973 |
| Rampage school Fatalities | GMAL | 28 | -.012 | .015 | -.816 | .415 | 9305 | .975 |
| Rampage school Fatalities | GMAL | 24 | -.01 | .01 | -1.026 | .305 | 12408 | .91 |
| Rampage school Fatalities | GMAL | 20 | 0 | .008 | -.034 | .973 | 15511 | .909 |
| Rampage school Fatalities | GMAL | 16 | .012 | .006 | 1.874 | .061 | 18614 | .91 |
| Rampage school Fatalities | GMAL | 12 | 0 | .006 | -.008 | .994 | 18616 | .945 |
| Rampage school Fatalities | GMAL | 8 | 0 | .006 | .059 | .953 | 18618 | .942 |
| Rampage school Fatalities | GMAL | 4 | -.009 | .004 | -2.162 | .031 | 18619 | .973 |
| Rampage school shootings | GMAL | 28 | -.015 | .017 | -.885 | .376 | 9305 | .975 |
| Rampage school shootings | GMAL | 24 | .012 | .013 | .91 | .363 | 12408 | .91 |
| Rampage school shootings | GMAL | 20 | .011 | .007 | 1.52 | .129 | 15511 | .909 |
| Rampage school shootings | GMAL | 16 | .001 | .009 | .107 | .915 | 18614 | .91 |
| Rampage school shootings | GMAL | 12 | .002 | .007 | .333 | .739 | 18616 | .945 |
| Rampage school shootings | GMAL | 8 | .004 | .008 | .458 | .647 | 18618 | .942 |
| Rampage school shootings | GMAL | 4 | .001 | .007 | .101 | .92 | 18619 | .973 |

Table S16: Figure 3e Results

| treatment | study | Years Prior | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|---|---|
| All mass shootings | YOUSAF | 8 | .009 | .008 | 1.196 | .232 | 9270 | .991 |
| All mass shootings | YOUSAF | 4 | -.004 | .007 | -.522 | .601 | 12432 | .912 |
| Rampage school No Fatalities | GMAL | 28 | -.007 | .022 | -.305 | .76 | 9305 | .975 |
| Rampage school No Fatalities | GMAL | 24 | .015 | .014 | 1.095 | .274 | 12408 | .91 |
| Rampage school No Fatalities | GMAL | 20 | .02 | .01 | 2.04 | .041 | 15511 | .909 |
| Rampage school No Fatalities | GMAL | 16 | -.002 | .009 | -.196 | .845 | 18614 | .91 |
| Rampage school No Fatalities | GMAL | 12 | -.004 | .007 | -.578 | .563 | 18616 | .945 |
| Rampage school No Fatalities | GMAL | 8 | -.007 | .008 | -.99 | .322 | 18618 | .942 |
| Rampage school No Fatalities | GMAL | 4 | .008 | .007 | 1.194 | .233 | 18619 | .973 |
| Rampage school Fatalities | GMAL | 28 | -.015 | .022 | -.705 | .481 | 9305 | .975 |
| Rampage school Fatalities | GMAL | 24 | -.005 | .013 | -.394 | .694 | 12408 | .91 |
| Rampage school Fatalities | GMAL | 20 | .001 | .01 | .13 | .897 | 15511 | .909 |
| Rampage school Fatalities | GMAL | 16 | .009 | .006 | 1.448 | .148 | 18614 | .91 |
| Rampage school Fatalities | GMAL | 12 | -.001 | .008 | -.071 | .944 | 18616 | .945 |
| Rampage school Fatalities | GMAL | 8 | .001 | .008 | .126 | .9 | 18618 | .942 |
| Rampage school Fatalities | GMAL | 4 | -.009 | .005 | -1.752 | .08 | 18619 | .973 |
| Rampage school shootings | GMAL | 28 | -.011 | .015 | -.727 | .467 | 9305 | .975 |
| Rampage school shootings | GMAL | 24 | .004 | .009 | .438 | .661 | 12408 | .91 |
| Rampage school shootings | GMAL | 20 | .01 | .008 | 1.254 | .21 | 15511 | .909 |
| Rampage school shootings | GMAL | 16 | .005 | .006 | .932 | .352 | 18614 | .91 |
| Rampage school shootings | GMAL | 12 | -.002 | .006 | -.365 | .715 | 18616 | .945 |
| Rampage school shootings | GMAL | 8 | -.003 | .006 | -.455 | .649 | 18618 | .942 |
| Rampage school shootings | GMAL | 4 | -.002 | .004 | -.478 | .633 | 18619 | .973 |

Table S17: Figure 3f Results

| treatment | study | Years Prior | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|---|---|
| All mass shootings | YOUSAF | 8 | .002 | .008 | .182 | .855 | 9270 | .991 |
| All mass shootings | YOUSAF | 4 | -.005 | .011 | -.45 | .653 | 12432 | .912 |
| Rampage school No Fatalities | GMAL | 28 | -.009 | .02 | -.427 | .669 | 9305 | .975 |
| Rampage school No Fatalities | GMAL | 24 | .012 | .014 | .836 | .403 | 12408 | .91 |
| Rampage school No Fatalities | GMAL | 20 | .02 | .009 | 2.124 | .034 | 15511 | .909 |
| Rampage school No Fatalities | GMAL | 16 | 0 | .008 | .05 | .96 | 18614 | .91 |
| Rampage school No Fatalities | GMAL | 12 | -.004 | .007 | -.623 | .534 | 18616 | .945 |
| Rampage school No Fatalities | GMAL | 8 | -.006 | .008 | -.777 | .437 | 18618 | .942 |
| Rampage school No Fatalities | GMAL | 4 | .005 | .007 | .796 | .426 | 18619 | .973 |
| Rampage school Fatalities | GMAL | 28 | -.005 | .02 | -.271 | .786 | 9305 | .975 |
| Rampage school Fatalities | GMAL | 24 | -.01 | .015 | -.68 | .496 | 12408 | .91 |
| Rampage school Fatalities | GMAL | 20 | .003 | .009 | .345 | .73 | 15511 | .909 |
| Rampage school Fatalities | GMAL | 16 | .008 | .006 | 1.179 | .239 | 18614 | .91 |
| Rampage school Fatalities | GMAL | 12 | .001 | .007 | .185 | .853 | 18616 | .945 |
| Rampage school Fatalities | GMAL | 8 | .004 | .008 | .542 | .588 | 18618 | .942 |
| Rampage school Fatalities | GMAL | 4 | -.009 | .005 | -1.835 | .067 | 18619 | .973 |
| Rampage school shootings | GMAL | 28 | -.015 | .023 | -.622 | .534 | 9305 | .975 |
| Rampage school shootings | GMAL | 24 | .012 | .016 | .744 | .457 | 12408 | .91 |
| Rampage school shootings | GMAL | 20 | .011 | .009 | 1.316 | .188 | 15511 | .909 |
| Rampage school shootings | GMAL | 16 | .001 | .01 | .089 | .929 | 18614 | .91 |
| Rampage school shootings | GMAL | 12 | .002 | .007 | .281 | .779 | 18616 | .945 |
| Rampage school shootings | GMAL | 8 | .004 | .009 | .4 | .69 | 18618 | .942 |
| Rampage school shootings | GMAL | 4 | .001 | .007 | .082 | .935 | 18619 | .973 |

Table S18: Figure 4 Results

| Var | Coef. | Std. Err. | t | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| lead9 | -0.146 | 0.023 | -6.350 | 0.000 | -0.190 | -0.101 |
| lead8 | -0.132 | 0.017 | -7.770 | 0.000 | -0.166 | -0.099 |
| lead7 | -0.126 | 0.013 | -9.800 | 0.000 | -0.151 | -0.101 |
| lead6 | -0.102 | 0.013 | -8.030 | 0.000 | -0.127 | -0.077 |
| lead5 | -0.068 | 0.013 | -5.360 | 0.000 | -0.092 | -0.043 |
| lead4 | -0.053 | 0.010 | -5.070 | 0.000 | -0.073 | -0.032 |
| lead3 | -0.040 | 0.008 | -4.930 | 0.000 | -0.056 | -0.024 |
| lead2 | -0.008 | 0.006 | -1.330 | 0.182 | -0.021 | 0.004 |
| lag0 | 0.030 | 0.004 | 7.340 | 0.000 | 0.022 | 0.038 |
| lag1 | 0.037 | 0.008 | 4.780 | 0.000 | 0.022 | 0.052 |
| lag2 | 0.041 | 0.013 | 3.180 | 0.001 | 0.016 | 0.066 |
| lag3 | 0.048 | 0.018 | 2.710 | 0.007 | 0.013 | 0.083 |
| lag4 | 0.069 | 0.019 | 3.530 | 0.000 | 0.030 | 0.107 |
| lag5 | 0.101 | 0.020 | 4.930 | 0.000 | 0.061 | 0.141 |
| lag6 | 0.117 | 0.023 | 5.130 | 0.000 | 0.072 | 0.161 |
| lag7 | 0.130 | 0.028 | 4.710 | 0.000 | 0.076 | 0.184 |

Table S19: Figure 5a Results

| treatment | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|
| All school shootings | .001 | .01 | .117 | .907 | 30625 | .784 |
| All mass shootings | -.002 | .007 | -.289 | .773 | 15570 | .871 |
| Rampage school shootings w No Fatalities | .002 | .006 | .365 | .715 | 18620 | .967 |
| Rampage school shootings w Fatalities | .01 | .006 | 1.709 | .088 | 18620 | .968 |
| Rampage school shootings | .007 | .004 | 1.772 | .076 | 18620 | .967 |

Table S20: Figure 5b Results

| treatment | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|
| All school shootings | .001 | .011 | .109 | .913 | 30625 | .784 |
| All mass shootings | -.002 | .008 | -.253 | .8 | 15570 | .871 |
| Rampage school shootings w No Fatalities | .002 | .006 | .321 | .748 | 18620 | .967 |
| Rampage school shootings w Fatalities | .01 | .007 | 1.524 | .128 | 18620 | .967 |
| Rampage school shootings | .007 | .005 | 1.578 | .115 | 18620 | .967 |

Table S21: Figure 5c Results

| treatment | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|
| All school shootings | .008 | .017 | .456 | .649 | 27354 | .196 |
| All mass shootings | 0 | .013 | .011 | .991 | 12432 | .277 |
| Rampage school shootings w No Fatalities | -.006 | .008 | -.752 | .452 | 18619 | .689 |
| Rampage school shootings w Fatalities | .019 | .008 | 2.424 | .015 | 18619 | .689 |
| Rampage school shootings | .01 | .006 | 1.569 | .117 | 18619 | .689 |

Table S22: Figure 5d Results

| treatment | coef | stderr | tstat | pval | N | r2 |
|---|---|---|---|---|---|---|
| All school shootings | .008 | .018 | .426 | .67 | 27354 | .196 |
| All mass shootings | 0 | .017 | .008 | .993 | 12432 | .277 |
| Rampage school shootings w No Fatalities | -.006 | .009 | -.674 | .501 | 18619 | .689 |
| Rampage school shootings w Fatalities | .019 | .009 | 2.167 | .03 | 18619 | .689 |
| Rampage school shootings | .01 | .007 | 1.402 | .161 | 18619 | .689 |

Table S23: Figure 6a Results

| Var | Coef. | Std. Err. | t | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| _k_eq_m6 | -0.01 | 0.01 | -1.45 | 0.15 | -0.02 | 0.00 |
| _k_eq_m5 | -0.01 | 0.01 | -0.87 | 0.38 | -0.02 | 0.01 |
| _k_eq_m4 | 0.00 | 0.01 | -0.75 | 0.45 | -0.02 | 0.01 |
| _k_eq_m3 | -0.01 | 0.01 | -1.66 | 0.10 | -0.02 | 0.00 |
| _k_eq_m2 | 0.00 | 0.00 | -0.91 | 0.36 | -0.01 | 0.00 |
| _k_eq_p0 | 0.01 | 0.00 | 1.85 | 0.07 | 0.00 | 0.02 |
| _k_eq_p1 | 0.01 | 0.01 | 1.28 | 0.20 | -0.01 | 0.03 |
| _k_eq_p2 | 0.01 | 0.01 | 0.74 | 0.46 | -0.01 | 0.03 |
| _k_eq_p3 | 0.00 | 0.01 | 0.28 | 0.78 | -0.02 | 0.02 |
| _k_eq_p4 | 0.01 | 0.01 | 1.01 | 0.31 | -0.01 | 0.02 |
| _k_eq_p5 | 0.01 | 0.01 | 1.81 | 0.07 | 0.00 | 0.03 |
| _k_eq_p6 | 0.03 | 0.02 | 1.68 | 0.09 | -0.01 | 0.07 |

Table S24: Figure 6b Results

| Var | Coef. | Std. Err. | t | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| _k_eq_m6 | -.0113595 | .0086608 | -1.31 | 0.190 | -.0283409 | .005622 |
| _k_eq_m5 | -.0130866 | .0100614 | -1.30 | 0.193 | -.0328143 | .0066411 |
| _k_eq_m4 | -.0069502 | .0083399 | -0.83 | 0.405 | -.0233024 | .009402 |
| _k_eq_m3 | -.0096858 | .0067104 | -1.44 | 0.149 | -.0228431 | .0034714 |
| _k_eq_m2 | -.00139 | .0041335 | -0.34 | 0.737 | -.0094947 | .0067147 |
| _k_eq_p0 | .0101121 | .0054591 | 1.85 | 0.064 | -.0005917 | .020816 |
| _k_eq_p1 | .0099699 | .0100398 | 0.99 | 0.321 | -.0097154 | .0296552 |
| _k_eq_p2 | .0057723 | .0132517 | 0.44 | 0.663 | -.0202107 | .0317553 |
| _k_eq_p3 | -.0033458 | .0127822 | -0.26 | 0.794 | -.0284081 | .0217165 |
| _k_eq_p4 | -.0018293 | .0101831 | -0.18 | 0.857 | -.0217957 | .018137 |
| _k_eq_p5 | .0012657 | .0113552 | 0.11 | 0.911 | -.0209987 | .0235302 |
| _k_eq_p6 | .0065909 | .0238612 | 0.28 | 0.782 | -.0401945 | .0533763 |

Table S25: Figure 6c Results

| Var | Coef. | Std. Err. | t | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| _k_eq_m6 | -.0086642 | .0099251 | -0.87 | 0.383 | -.0281245 | .0107961 |
| _k_eq_m5 | .0008727 | .010344 | 0.08 | 0.933 | -.0194091 | .0211545 |
| _k_eq_m4 | -.003458 | .008579 | -0.40 | 0.687 | -.0202791 | .0133632 |
| _k_eq_m3 | -.0086557 | .00732 | -1.18 | 0.237 | -.0230082 | .0056968 |
| _k_eq_m2 | -.0088006 | .0066084 | -1.33 | 0.183 | -.0217579 | .0041567 |
| _k_eq_p0 | .0045428 | .0076024 | 0.60 | 0.550 | -.0103634 | .019449 |
| _k_eq_p1 | .0154326 | .015197 | 1.02 | 0.310 | -.0143646 | .0452298 |
| _k_eq_p2 | .0202693 | .0193697 | 1.05 | 0.295 | -.0177094 | .058248 |
| _k_eq_p3 | .0339329 | .0225165 | 1.51 | 0.132 | -.0102158 | .0780815 |
| _k_eq_p4 | .0547016 | .0336609 | 1.63 | 0.104 | -.0112983 | .1207014 |
| _k_eq_p5 | .0748566 | .0397417 | 1.88 | 0.060 | -.0030661 | .1527792 |
| _k_eq_p6 | .076861 | .0480658 | 1.60 | 0.110 | -.017383 | .1711049 |

Table S26: Figure 6d Results

| Var | Coef. | Std. Err. | t | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| _k_eq_m6 | .0042553 | .0171401 | 0.25 | 0.804 | -.0293517 | .0378622 |
| _k_eq_m5 | .0032087 | .0168424 | 0.19 | 0.849 | -.0298145 | .036232 |
| _k_eq_m4 | .0000179 | .0145303 | 0.00 | 0.999 | -.0284721 | .0285079 |
| _k_eq_m3 | .0092445 | .0131328 | 0.70 | 0.482 | -.0165053 | .0349943 |
| _k_eq_m2 | -.0000906 | .0108249 | -0.01 | 0.993 | -.0213152 | .0211341 |
| _k_eq_p0 | .0036256 | .0111273 | 0.33 | 0.745 | -.018192 | .0254433 |
| _k_eq_p1 | .0163643 | .0175472 | 0.93 | 0.351 | -.0180409 | .0507695 |
| _k_eq_p2 | .0082639 | .0235333 | 0.35 | 0.725 | -.0378785 | .0544064 |
| _k_eq_p3 | .0118279 | .0282368 | 0.42 | 0.675 | -.0435367 | .0671925 |
| _k_eq_p4 | .0080626 | .0281694 | 0.29 | 0.775 | -.0471699 | .0632952 |
| _k_eq_p5 | .0123471 | .0269939 | 0.46 | 0.647 | -.0405805 | .0652747 |
| _k_eq_p6 | -.0076323 | .0350421 | -0.22 | 0.828 | -.0763403 | .0610756 |

Table S27: Figure 7a Results

| Time | N | ATT | ATT_sd | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| -10 | 85 | -0.016 | 0.017 | 0.336 | -0.050 | 0.017 |
| -9 | 85 | 0.010 | 0.013 | 0.432 | -0.017 | 0.034 |
| -8 | 86 | -0.003 | 0.015 | 0.856 | -0.034 | 0.025 |
| -7 | 89 | 0.027 | 0.011 | 0.017 | 0.002 | 0.048 |
| -6 | 92 | -0.002 | 0.015 | 0.914 | -0.030 | 0.024 |
| -5 | 95 | 0.046 | 0.011 | 0.000 | 0.024 | 0.068 |
| -4 | 100 | 0.002 | 0.019 | 0.908 | -0.038 | 0.038 |
| -3 | 103 | 0.010 | 0.015 | 0.521 | -0.022 | 0.039 |
| -2 | 106 | 0.023 | 0.017 | 0.170 | -0.005 | 0.058 |
| -1 | 110 | 0.038 | 0.014 | 0.007 | 0.012 | 0.068 |
| 0 | 110 | 0.022 | 0.022 | 0.309 | -0.025 | 0.061 |
| 1 | 111 | 0.043 | 0.019 | 0.021 | 0.008 | 0.079 |
| 2 | 104 | 0.046 | 0.020 | 0.018 | 0.009 | 0.083 |
| 3 | 97 | 0.075 | 0.016 | 0.000 | 0.045 | 0.109 |
| 4 | 86 | 0.046 | 0.024 | 0.052 | -0.003 | 0.090 |

Table S28: Figure 7b Results

| Time | N | ATT | ATT_sd | p | LB95 | UB |
|---|---|---|---|---|---|---|
| -10 | 85 | -0.016 | 0.017 | 0.336 | -0.050 | 0.017 |
| -9 | 85 | 0.010 | 0.013 | 0.432 | -0.017 | 0.034 |
| -8 | 86 | -0.003 | 0.015 | 0.856 | -0.034 | 0.025 |
| -7 | 89 | 0.027 | 0.011 | 0.017 | 0.002 | 0.048 |
| -6 | 92 | -0.002 | 0.015 | 0.914 | -0.030 | 0.024 |
| -5 | 95 | 0.046 | 0.011 | 0.000 | 0.024 | 0.068 |
| -4 | 100 | 0.002 | 0.019 | 0.908 | -0.038 | 0.038 |
| -3 | 103 | 0.010 | 0.015 | 0.521 | -0.022 | 0.039 |
| -2 | 106 | 0.023 | 0.017 | 0.170 | -0.005 | 0.058 |
| -1 | 110 | 0.038 | 0.014 | 0.007 | 0.012 | 0.068 |
| 0 | 110 | 0.022 | 0.022 | 0.309 | -0.025 | 0.061 |
| 1 | 111 | 0.043 | 0.019 | 0.021 | 0.008 | 0.079 |
| 2 | 104 | 0.046 | 0.020 | 0.018 | 0.009 | 0.083 |
| 3 | 97 | 0.075 | 0.016 | 0.000 | 0.045 | 0.109 |
| 4 | 86 | 0.046 | 0.024 | 0.052 | -0.003 | 0.090 |

Table S29: Figure 7c Results

| Time | N | ATT | ATT_sd | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| -10 | 85 | -0.011 | 0.006 | 0.103 | -0.024 | 0.002 |
| -9 | 85 | -0.001 | 0.009 | 0.918 | -0.017 | 0.016 |
| -8 | 86 | -0.002 | 0.010 | 0.873 | -0.020 | 0.018 |
| -7 | 89 | 0.012 | 0.006 | 0.057 | 0.001 | 0.026 |
| -6 | 92 | -0.012 | 0.006 | 0.047 | -0.024 | -0.001 |
| -5 | 95 | 0.004 | 0.006 | 0.530 | -0.006 | 0.017 |
| -4 | 100 | -0.001 | 0.007 | 0.848 | -0.017 | 0.010 |
| -3 | 103 | -0.010 | 0.005 | 0.032 | -0.019 | -0.001 |
| -2 | 106 | 0.003 | 0.006 | 0.632 | -0.008 | 0.014 |
| -1 | 110 | 0.001 | 0.006 | 0.885 | -0.010 | 0.012 |
| 0 | 110 | -0.007 | 0.008 | 0.394 | -0.021 | 0.011 |
| 1 | 111 | -0.003 | 0.008 | 0.738 | -0.018 | 0.013 |
| 2 | 104 | 0.013 | 0.010 | 0.203 | -0.006 | 0.036 |
| 3 | 97 | 0.015 | 0.008 | 0.047 | 0.000 | 0.031 |
| 4 | 86 | 0.010 | 0.009 | 0.276 | -0.010 | 0.026 |

Table S30: Figure 7d Results

| Time | N | ATT | ATT_sd | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| -10 | 85 | -0.011 | 0.007 | 0.115 | -0.023 | 0.003 |
| -9 | 85 | -0.001 | 0.009 | 0.916 | -0.023 | 0.015 |
| -8 | 86 | -0.002 | 0.010 | 0.881 | -0.019 | 0.022 |
| -7 | 89 | 0.012 | 0.006 | 0.044 | 0.001 | 0.024 |
| -6 | 92 | -0.012 | 0.005 | 0.032 | -0.021 | 0.000 |
| -5 | 95 | 0.004 | 0.006 | 0.534 | -0.009 | 0.017 |
| -4 | 100 | -0.001 | 0.006 | 0.822 | -0.014 | 0.009 |
| -3 | 103 | -0.010 | 0.005 | 0.035 | -0.020 | -0.001 |
| -2 | 106 | 0.003 | 0.005 | 0.607 | -0.008 | 0.013 |
| -1 | 110 | 0.001 | 0.006 | 0.888 | -0.012 | 0.014 |
| 0 | 110 | -0.007 | 0.008 | 0.427 | -0.027 | 0.006 |
| 1 | 111 | -0.003 | 0.007 | 0.721 | -0.017 | 0.013 |
| 2 | 104 | 0.013 | 0.009 | 0.173 | -0.005 | 0.030 |
| 3 | 97 | 0.015 | 0.008 | 0.061 | 0.000 | 0.031 |
| 4 | 86 | 0.010 | 0.009 | 0.261 | -0.006 | 0.027 |

Table S31: Figure 8a Results

| lb | ub | M |
|---|---|---|
| 0.98 | 7.88 | Original |
| -2.15 | 4.66 | 0 |
| -2.2 | 4.7 | 0.02 |
| -2.35 | 4.86 | 0.04 |
| -2.58 | 5.08 | 0.06 |
| -2.85 | 5.36 | 0.08 |
| -2.47 | 6.32 | 0.1 |

Table S32: Figure 8b Results

| lb | ub | M |
|---|---|---|
| 0.159 | 0.233 | Original |
| -1 | 1 | 0 |
| -1 | 1 | 0.5 |
| -1 | 1 | 1 |
| -1 | 1 | 1.5 |
| -1 | 1 | 2 |

Table S33: Figure 9a Results

| Bacon_treat | Bacon_control | Bacon_weight | Bacon_coef | Bacon_cgroup |
|---|---|---|---|---|
| 1984 | 1988 | 8.70E-08 | 0.078856 | Early vs Late |
| 1988 | 1984 | 7.00E-07 | -0.0695 | Late vs Early |
| 1984 | 1992 | 4.40E-08 | -0.01871 | Early vs Late |
| 1992 | 1984 | 3.10E-07 | -0.081747 | Late vs Early |
| 1988 | 1992 | 7.50E-08 | -0.036729 | Early vs Late |
| 1992 | 1988 | 2.60E-07 | 0.006331 | Late vs Early |
| 1984 | 1996 | 1.30E-07 | 0.027982 | Early vs Late |
| 1996 | 1984 | 7.90E-07 | -0.06679 | Late vs Early |
| 1988 | 1996 | 3.00E-07 | -0.037058 | Early vs Late |
| 1996 | 1988 | 9.00E-07 | 0.016298 | Late vs Early |
| 1992 | 1996 | 5.60E-08 | -0.028406 | Early vs Late |
| 1996 | 1992 | 1.10E-07 | -0.020474 | Late vs Early |
| 1984 | 2000 | 2.90E-07 | 0.098948 | Early vs Late |
| 2000 | 1984 | 1.50E-06 | -0.126772 | Late vs Early |
| 1988 | 2000 | 7.50E-07 | 0.019507 | Early vs Late |
| 2000 | 1988 | 1.90E-06 | -0.022988 | Late vs Early |
| 1992 | 2000 | 1.90E-07 | 0.009956 | Early vs Late |
| 2000 | 1992 | 3.10E-07 | -0.070383 | Late vs Early |
| 1996 | 2000 | 2.50E-07 | 0.044571 | Early vs Late |
| 2000 | 1996 | 3.10E-07 | -0.032647 | Late vs Early |
| 1984 | 2004 | 3.30E-07 | 0.017506 | Early vs Late |
| 2004 | 1984 | 1.30E-06 | -0.07576 | Late vs Early |
| 1988 | 2004 | 9.00E-07 | -0.033137 | Early vs Late |
| 2004 | 1988 | 1.80E-06 | 0.023813 | Late vs Early |
| 1992 | 2004 | 2.50E-07 | -0.028567 | Early vs Late |
| 2004 | 1992 | 3.40E-07 | -0.031244 | Late vs Early |
| 1996 | 2004 | 4.50E-07 | -0.006315 | Early vs Late |
| 2004 | 1996 | 4.50E-07 | -0.002225 | Late vs Early |
| 2000 | 2004 | 4.70E-07 | -0.052952 | Early vs Late |
| 2004 | 2000 | 3.70E-07 | 0.025412 | Late vs Early |
| 1984 | 2008 | 5.20E-07 | 0.046045 | Early vs Late |
| 2008 | 1984 | 1.60E-06 | -0.04371 | Late vs Early |
| 1988 | 2008 | 1.50E-06 | -0.034392 | Early vs Late |
| 2008 | 1988 | 2.20E-06 | 0.051818 | Late vs Early |
| 1992 | 2008 | 4.50E-07 | -0.025771 | Early vs Late |
| 2008 | 1992 | 4.50E-07 | -0.012751 | Late vs Early |
| 1996 | 2008 | 9.00E-07 | 0.008181 | Early vs Late |
| 2008 | 1996 | 6.70E-07 | 0.042243 | Late vs Early |
| 2000 | 2008 | 1.20E-06 | -0.048998 | Early vs Late |
| 2008 | 2000 | 7.50E-07 | 0.062911 | Late vs Early |
| 2004 | 2008 | 6.70E-07 | -0.005337 | Early vs Late |

Table S33: Figure 9a Results

| Bacon_treat | Bacon_control | Bacon_weight | Bacon_coef | Bacon_cgroup |
|---|---|---|---|---|
| 2008 | 2004 | 3.40E-07 | 0.030546 | Late vs Early |
| 1984 | 2012 | 6.60E-07 | 0.03054 | Early vs Late |
| 2012 | 1984 | 1.30E-06 | -0.022081 | Late vs Early |
| 1988 | 2012 | 1.90E-06 | -0.059704 | Early vs Late |
| 2012 | 1988 | 1.90E-06 | 0.071537 | Late vs Early |
| 1992 | 2012 | 6.10E-07 | -0.04026 | Early vs Late |
| 2012 | 1992 | 4.10E-07 | 0.005722 | Late vs Early |
| 1996 | 2012 | 1.30E-06 | -0.018156 | Early vs Late |
| 2012 | 1996 | 6.50E-07 | 0.053814 | Late vs Early |
| 2000 | 2012 | 2.00E-06 | -0.081882 | Early vs Late |
| 2012 | 2000 | 8.10E-07 | 0.062162 | Late vs Early |
| 2004 | 2012 | 1.50E-06 | -0.021788 | Early vs Late |
| 2012 | 2004 | 4.90E-07 | 0.046749 | Late vs Early |
| 2008 | 2012 | 1.10E-06 | -0.005292 | Early vs Late |
| 2012 | 2008 | 3.20E-07 | 0.024165 | Late vs Early |
| 1984 | 2016 | 1.50E-06 | 0.06668 | Early vs Late |
| 2016 | 1984 | 1.50E-06 | -0.031652 | Late vs Early |
| 1988 | 2016 | 4.50E-06 | -0.044903 | Early vs Late |
| 2016 | 1988 | 2.30E-06 | 0.060484 | Late vs Early |
| 1992 | 2016 | 1.50E-06 | -0.02972 | Early vs Late |
| 2016 | 1992 | 4.90E-07 | -0.006247 | Late vs Early |
| 1996 | 2016 | 3.20E-06 | -0.017525 | Early vs Late |
| 2016 | 1996 | 8.10E-07 | 0.040247 | Late vs Early |
| 2000 | 2016 | 5.40E-06 | -0.084005 | Early vs Late |
| 2016 | 2000 | 1.10E-06 | 0.05345 | Late vs Early |
| 2004 | 2016 | 4.40E-06 | -0.016841 | Early vs Late |
| 2016 | 2004 | 7.30E-07 | 0.042292 | Late vs Early |
| 2008 | 2016 | 4.50E-06 | 0.007404 | Early vs Late |
| 2016 | 2008 | 6.50E-07 | 0.017441 | Late vs Early |
| 2012 | 2016 | 2.80E-06 | 0.018737 | Early vs Late |
| 2016 | 2012 | 3.50E-07 | -0.012987 | Late vs Early |
| 1984 | Always | 6.60E-08 | 0.178325 | Always treated vs timing |
| 1988 | Always | 2.00E-07 | 0.019616 | Always treated vs timing |
| 1992 | Always | 6.60E-08 | 0.052281 | Always treated vs timing |
| 1996 | Always | 1.50E-07 | 0.100607 | Always treated vs timing |
| 2000 | Always | 2.60E-07 | 0.080886 | Always treated vs timing |
| 2004 | Always | 2.20E-07 | 0.170976 | Always treated vs timing |
| 2008 | Always | 2.60E-07 | 0.219987 | Always treated vs timing |
| 2012 | Always | 2.20E-07 | 0.207109 | Always treated vs timing |
| 2016 | Always | 2.40E-07 | 0.175516 | Always treated vs timing |
| 1984 | Never | 0.000197 | 0.15999 | Never treated vs timing |

Table S33: Figure 9a Results

| Bacon_treat | Bacon_control | Bacon_weight | Bacon_coef | Bacon_cgroup |
|---|---|---|---|---|
| 1988 | Never | 0.0006 | 0.031788 | Never treated vs timing |
| 1992 | Never | 0.000197 | 0.070029 | Never treated vs timing |
| 1996 | Never | 0.00045 | 0.081049 | Never treated vs timing |
| 2000 | Never | 0.000781 | 0.017108 | Never treated vs timing |
| 2004 | Never | 0.000675 | 0.084976 | Never treated vs timing |
| 2008 | Never | 0.000787 | 0.114054 | Never treated vs timing |
| 2012 | Never | 0.00065 | 0.138089 | Never treated vs timing |
| 2016 | Never | 0.000731 | 0.13174 | Never treated vs timing |

Table S34: Figure 9b Results

| Bacon_treat | Bacon_control | Bacon_weight | Bacon_coef | Bacon_cgroup |
|---|---|---|---|---|
| 2014 | 2012 | 5.60E-07 | -0.05964 | Late vs Early |
| 2012 | 2010 | 6.50E-08 | 0.003213 | Late vs Early |
| 2008 | 2006 | 4.40E-07 | 0.085913 | Late vs Early |
| 2014 | 2004 | 9.30E-07 | 0.022325 | Late vs Early |
| 2012 | 2008 | 3.90E-07 | 0.039837 | Late vs Early |
| 2010 | 2008 | 8.10E-08 | 0.092414 | Late vs Early |
| 2014 | 2008 | 1.70E-06 | -0.03245 | Late vs Early |
| 2008 | 2004 | 1.90E-07 | 0.107997 | Late vs Early |
| 2008 | 2002 | 7.30E-07 | 0.117028 | Late vs Early |
| 2014 | 2010 | 3.70E-07 | -0.052226 | Late vs Early |
| 2016 | 2008 | 4.50E-07 | -0.020182 | Late vs Early |
| 2016 | 2006 | 8.50E-07 | -0.043381 | Late vs Early |
| 2010 | 2004 | 8.10E-08 | 0.117657 | Late vs Early |
| 2016 | 2014 | 4.30E-07 | -0.020781 | Late vs Early |
| 2016 | 2004 | 2.30E-07 | -0.004238 | Late vs Early |
| 2016 | 2012 | 2.30E-07 | -0.074471 | Late vs Early |
| 2006 | 2004 | 1.70E-07 | -0.011611 | Late vs Early |
| 2018 | 2002 | 9.20E-07 | 0.031138 | Late vs Early |
| 2018 | 2012 | 4.10E-07 | -0.040066 | Late vs Early |
| 2018 | 2006 | 1.20E-06 | -0.011648 | Late vs Early |
| 2010 | 2002 | 2.70E-07 | 0.118213 | Late vs Early |
| 2014 | 2002 | 2.80E-06 | 0.029688 | Late vs Early |
| 2018 | 2014 | 1.10E-06 | 0.040257 | Late vs Early |
| 2018 | 2004 | 3.20E-07 | 0.025082 | Late vs Early |
| 2018 | 2016 | 1.60E-07 | 0.044396 | Late vs Early |
| 2016 | 2002 | 6.60E-07 | -0.022813 | Late vs Early |
| 2016 | 2010 | 1.10E-07 | -0.094541 | Late vs Early |
| 2012 | 2006 | 8.70E-07 | 0.018694 | Late vs Early |

Table S34: Figure 9b Results

| Bacon_treat | Bacon_control | Bacon_weight | Bacon_coef | Bacon_cgroup |
|---|---|---|---|---|
| 2006 | 2002 | 8.50E-07 | 0.056135 | Late vs Early |
| 2014 | 2006 | 3.30E-06 | -0.037409 | Late vs Early |
| 2004 | 2002 | 1.10E-07 | 0.089145 | Late vs Early |
| 2010 | 2006 | 2.40E-07 | 0.066528 | Late vs Early |
| 2012 | 2002 | 8.10E-07 | 0.10034 | Late vs Early |
| 2018 | 2008 | 6.90E-07 | -0.023922 | Late vs Early |
| 2012 | 2004 | 2.60E-07 | 0.068792 | Late vs Early |
| 2018 | 2010 | 1.80E-07 | -0.06747 | Late vs Early |
| 2006 | 2014 | 3.30E-06 | -0.017836 | Early vs Late |
| 2004 | 2014 | 6.20E-07 | 0.020437 | Early vs Late |
| 2014 | 2018 | 7.40E-06 | 0.018665 | Early vs Late |
| 2006 | 2012 | 6.50E-07 | -0.000757 | Early vs Late |
| 2002 | 2016 | 3.30E-07 | -0.083052 | Early vs Late |
| 2008 | 2014 | 2.20E-06 | 0.050286 | Early vs Late |
| 2008 | 2012 | 3.90E-07 | 0.080038 | Early vs Late |
| 2004 | 2018 | 6.40E-07 | 0.023406 | Early vs Late |
| 2006 | 2010 | 1.50E-07 | 0.075553 | Early vs Late |
| 2006 | 2016 | 1.30E-06 | -0.01 | Early vs Late |
| 2006 | 2018 | 3.70E-06 | 0.003769 | Early vs Late |
| 2002 | 2012 | 2.00E-07 | 0.049098 | Early vs Late |
| 2012 | 2016 | 6.80E-07 | 0.004417 | Early vs Late |
| 2016 | 2018 | 1.30E-06 | 0.019798 | Early vs Late |
| 2008 | 2016 | 9.10E-07 | 0.061464 | Early vs Late |
| 2010 | 2012 | 8.10E-08 | 0.004033 | Early vs Late |
| 2002 | 2010 | 5.40E-08 | 0.154867 | Early vs Late |
| 2002 | 2004 | 1.30E-08 | -0.003687 | Early vs Late |
| 2004 | 2012 | 1.30E-07 | 0.048794 | Early vs Late |
| 2002 | 2014 | 9.30E-07 | -0.0472 | Early vs Late |
| 2004 | 2006 | 4.90E-08 | -0.014637 | Early vs Late |
| 2002 | 2006 | 1.20E-07 | -0.013904 | Early vs Late |
| 2004 | 2016 | 2.30E-07 | 0.023891 | Early vs Late |
| 2012 | 2014 | 1.10E-06 | -0.006498 | Early vs Late |
| 2006 | 2008 | 2.20E-07 | 0.036324 | Early vs Late |
| 2010 | 2016 | 2.80E-07 | -0.01748 | Early vs Late |
| 2004 | 2008 | 6.50E-08 | 0.05269 | Early vs Late |
| 2012 | 2018 | 2.50E-06 | 0.031355 | Early vs Late |
| 2014 | 2016 | 1.50E-06 | -0.027395 | Early vs Late |
| 2010 | 2014 | 6.20E-07 | -0.004043 | Early vs Late |
| 2008 | 2018 | 2.70E-06 | 0.086759 | Early vs Late |
| 2004 | 2010 | 3.20E-08 | 0.189549 | Early vs Late |
| 2002 | 2008 | 1.20E-07 | -0.01369 | Early vs Late |

Table S34: Figure 9b Results

| Bacon_treat | Bacon_control | Bacon_weight | Bacon_coef | Bacon_cgroup |
|---|---|---|---|---|
| 2010 | 2018 | 9.20E-07 | 0.023001 | Early vs Late |
| 2002 | 2018 | 9.20E-07 | -0.041496 | Early vs Late |
| 2008 | 2010 | 6.50E-08 | 0.150448 | Early vs Late |
| 2014 | Always | 2.60E-06 | -0.097161 | Always treated vs timing |
| 2004 | Always | 1.70E-07 | 0.000735 | Always treated vs timing |
| 2010 | Always | 2.70E-07 | -0.060252 | Always treated vs timing |
| 2006 | Always | 1.00E-06 | -0.030167 | Always treated vs timing |
| 2012 | Always | 7.80E-07 | -0.065181 | Always treated vs timing |
| 2002 | Always | 2.40E-07 | -0.190752 | Always treated vs timing |
| 2018 | Always | 8.20E-07 | -0.030875 | Always treated vs timing |
| 2016 | Always | 6.00E-07 | -0.066575 | Always treated vs timing |
| 2008 | Always | 7.80E-07 | 0.022318 | Always treated vs timing |
| 2004 | Never | 0.000114 | 0.084458 | Never treated vs timing |
| 2010 | Never | 0.000178 | 0.094548 | Never treated vs timing |
| 2006 | Never | 0.000673 | 0.057976 | Never treated vs timing |
| 2016 | Never | 0.000399 | 0.067682 | Never treated vs timing |
| 2012 | Never | 0.000513 | 0.107867 | Never treated vs timing |
| 2014 | Never | 0.00172 | 0.063393 | Never treated vs timing |
| 2008 | Never | 0.000513 | 0.134249 | Never treated vs timing |
| 2018 | Never | 0.000545 | 0.068284 | Never treated vs timing |
| 2002 | Never | 0.00016 | 0.023084 | Never treated vs timing |

Table S35: Figure 10a Results

| Var | Coef. | Std. Err. | t | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| g_16 | 0.018 | 0.013 | 1.350 | 0.176 | -0.008 | 0.043 |
| g_12 | 0.007 | 0.010 | 0.730 | 0.465 | -0.012 | 0.027 |
| g_8 | -0.001 | 0.008 | -0.080 | 0.933 | -0.015 | 0.014 |
| g0 | 0.015 | 0.007 | 1.960 | 0.050 | 0.000 | 0.029 |
| g4 | 0.007 | 0.009 | 0.820 | 0.410 | -0.010 | 0.024 |
| g8 | 0.006 | 0.009 | 0.720 | 0.469 | -0.011 | 0.023 |
| g12 | 0.002 | 0.007 | 0.300 | 0.761 | -0.012 | 0.016 |
| g16 | 0.005 | 0.004 | 1.260 | 0.209 | -0.003 | 0.013 |

Table S36: Figure 10b Results

| Var | Coef. | Std. Err. | t | p | LB95 | UB95 |
|---|---|---|---|---|---|---|
| g_14 | -0.029 | 0.033 | -0.880 | 0.378 | -0.095 | 0.036 |
| g_12 | 0.002 | 0.034 | 0.050 | 0.959 | -0.065 | 0.069 |
| g_10 | -0.009 | 0.027 | -0.340 | 0.736 | -0.062 | 0.044 |
| g_8 | -0.004 | 0.024 | -0.150 | 0.880 | -0.051 | 0.044 |
| g_6 | -0.001 | 0.021 | -0.020 | 0.981 | -0.041 | 0.040 |
| g_4 | -0.016 | 0.018 | -0.910 | 0.365 | -0.052 | 0.019 |
| g0 | -0.030 | 0.023 | -1.320 | 0.187 | -0.075 | 0.015 |
| g2 | -0.017 | 0.019 | -0.910 | 0.365 | -0.055 | 0.020 |
| g4 | -0.037 | 0.019 | -1.950 | 0.051 | -0.074 | 0.000 |
| g6 | -0.046 | 0.031 | -1.470 | 0.141 | -0.106 | 0.015 |
| g8 | -0.071 | 0.035 | -2.030 | 0.042 | -0.139 | -0.002 |
| g10 | -0.039 | 0.029 | -1.320 | 0.188 | -0.096 | 0.019 |
| g12 | -0.042 | 0.024 | -1.730 | 0.083 | -0.089 | 0.005 |
| g14 | 0.020 | 0.023 | 0.870 | 0.385 | -0.025 | 0.065 |

Given Figure 2 in the manuscript, some may wonder if we discard never treated units and, instead, compare the treated units with the not-yet-but-eventually-treated units. Such could be valid comparison group. If they were, we could, perhaps, avoid taking a stand on the type of violations of parallel trends. Unfortunately, this is not the case. We still observe pre-treatmenet imbalances among this group. These are of similar magnitude to the effects observed post-treatment. Once trends are added, any evidence for an effect disappears. This is shown in the Table below. Though this approach doesn't work in ours, this comparison could be a viable option for applied researchers in other settings.

Table S37: Using Eventually Treated as the Control Group

| treatment | time | YearsPre | model | coef | tstat | stderr | pval | N | r2 |
|---|---|---|---|---|---|---|---|---|---|
| All school shootings | Post | -4 | Quad Trends | .006 | 1.378 | .004 | .171 | 990 | .946 |
| All school shootings | Post | -4 | Linear Trends | .006 | 1.471 | .004 | .144 | 990 | .946 |
| All school shootings | Post | -4 | TWFE | .018 | 2.452 | .007 | .016 | 990 | .794 |
| All school shootings | Pre | 20 | TWFE | .015 | 2.446 | .006 | .016 | 495 | .886 |
| All school shootings | Pre | 16 | TWFE | .009 | 1.585 | .005 | .116 | 594 | .875 |
| All school shootings | Pre | 12 | TWFE | .006 | .992 | .006 | .324 | 693 | .862 |
| All school shootings | Pre | 8 | TWFE | .019 | 2.728 | .007 | .008 | 792 | .837 |
| All school shootings | Pre | 4 | TWFE | .01 | 1.539 | .007 | .127 | 891 | .815 |