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

1 Commonalities in the Papers in Question

We think it important to note that all three papers in qustion have some common findings. GMAL, Yousaf, and HHB all find no effects of mass shootings on voter turnout (as we show below, turnout does not appear to have the trend differences that plague Democratic vote share—see Figure S10.) This overall finding is also corroborated by a recent paper by Marsh (2022), 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 Figure 1, Marsh 2022), 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 (2022), 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 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, but it provides a theoretical reason to question their results. 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 clear support for an effect on vote shares.

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

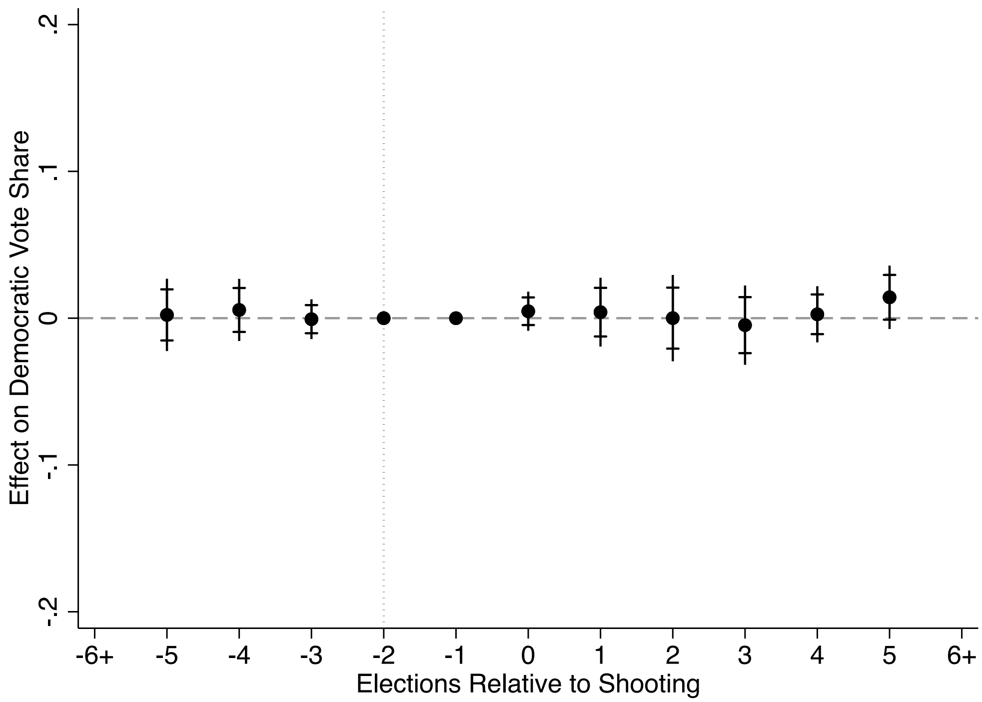
2 Solutions for Effect Heterogeneity Problems

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. 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.[1] 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*. 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 other methods afford and is more computationally intensive. Still, to illustrate this method, we provide the results for this in Tables S8 and S9. 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*. 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.

[1] Relevant packages here include the `csdid` in in *STATA* and in *R* and *hdidregress* and *xthdidregress* in *STATA* 18.

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

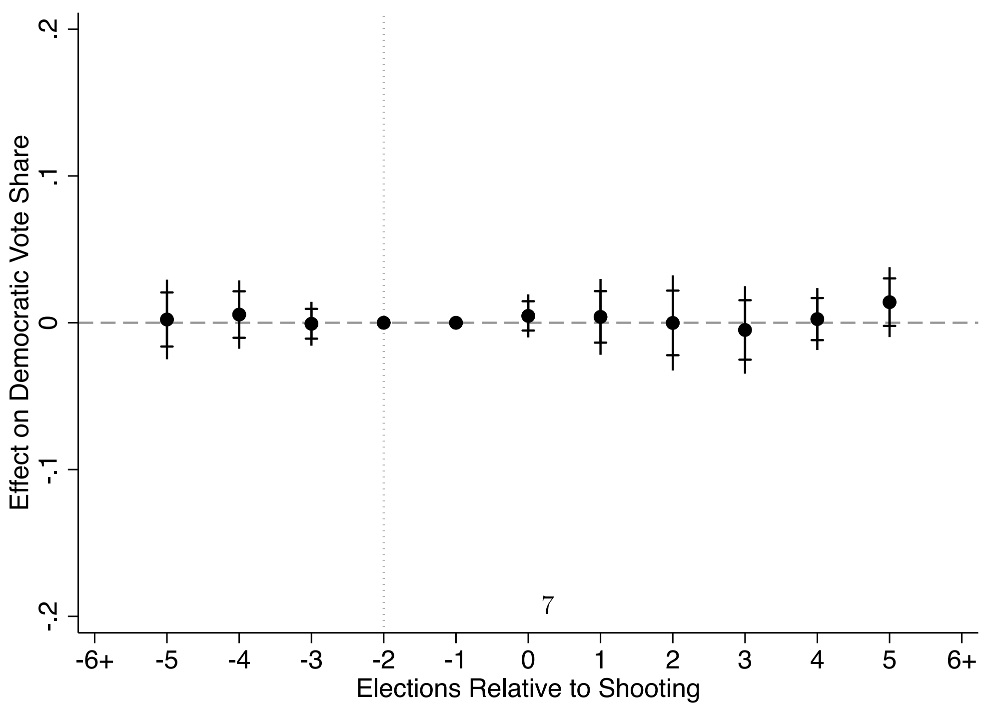
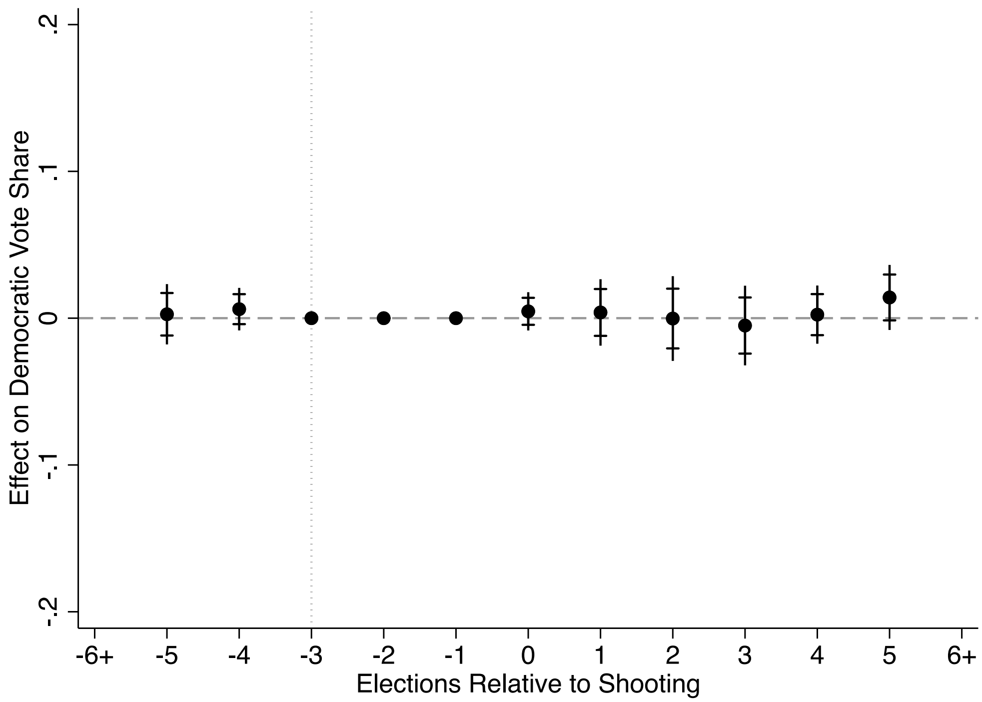


Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021). **Takeaway:** Benchmarked to pre-treatment trends at t-2, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

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

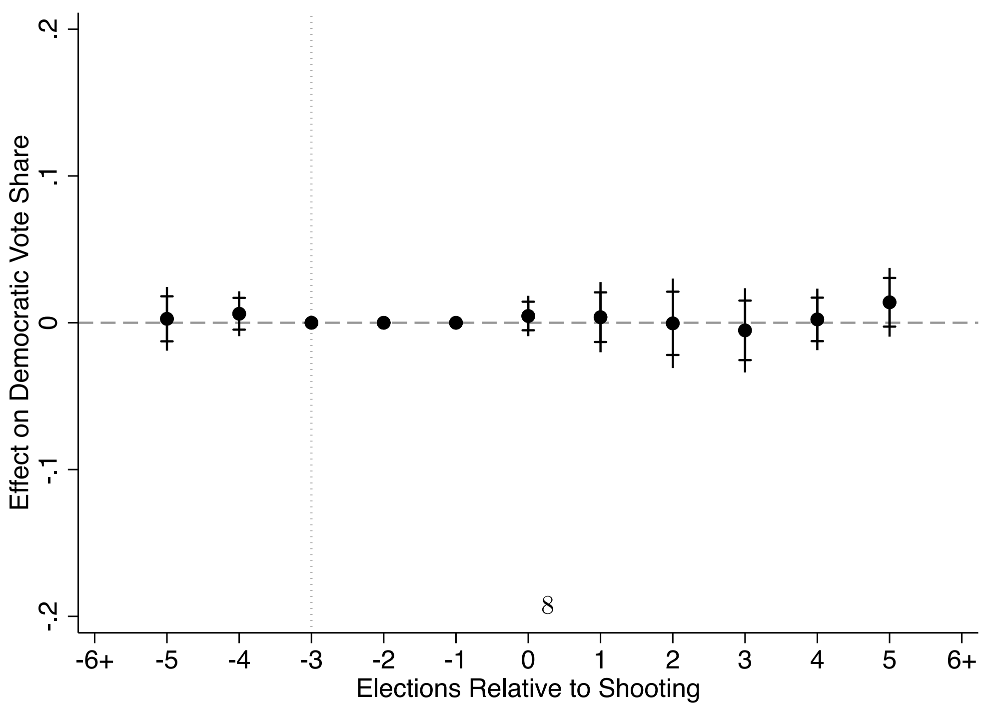
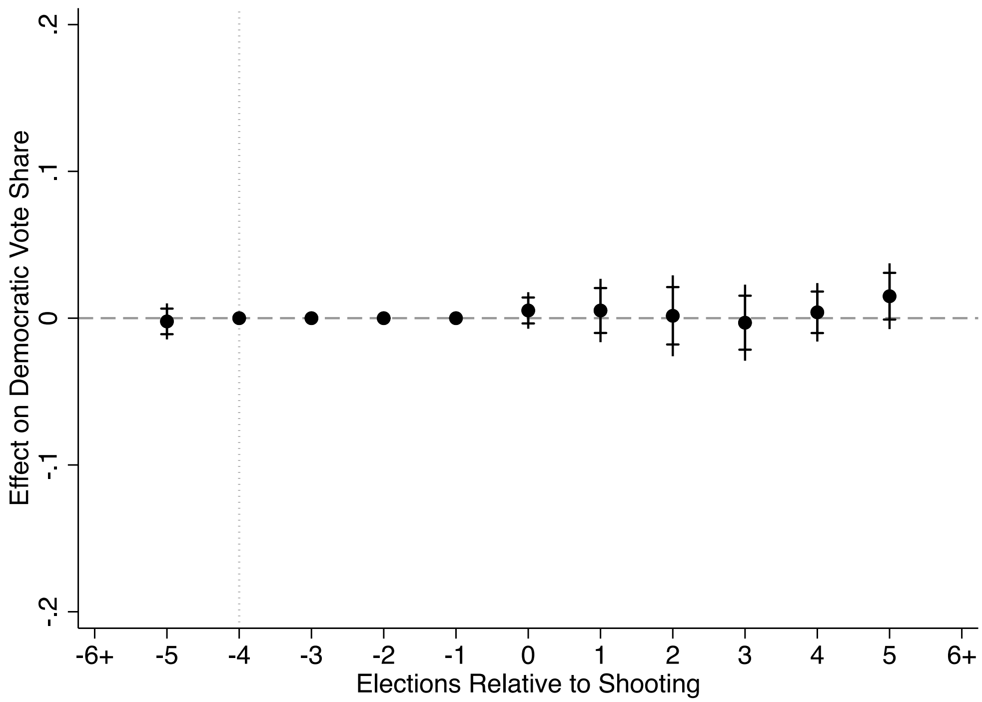


Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021). **Takeaway:** Benchmarked to pre-treatment trends at t-3, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

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

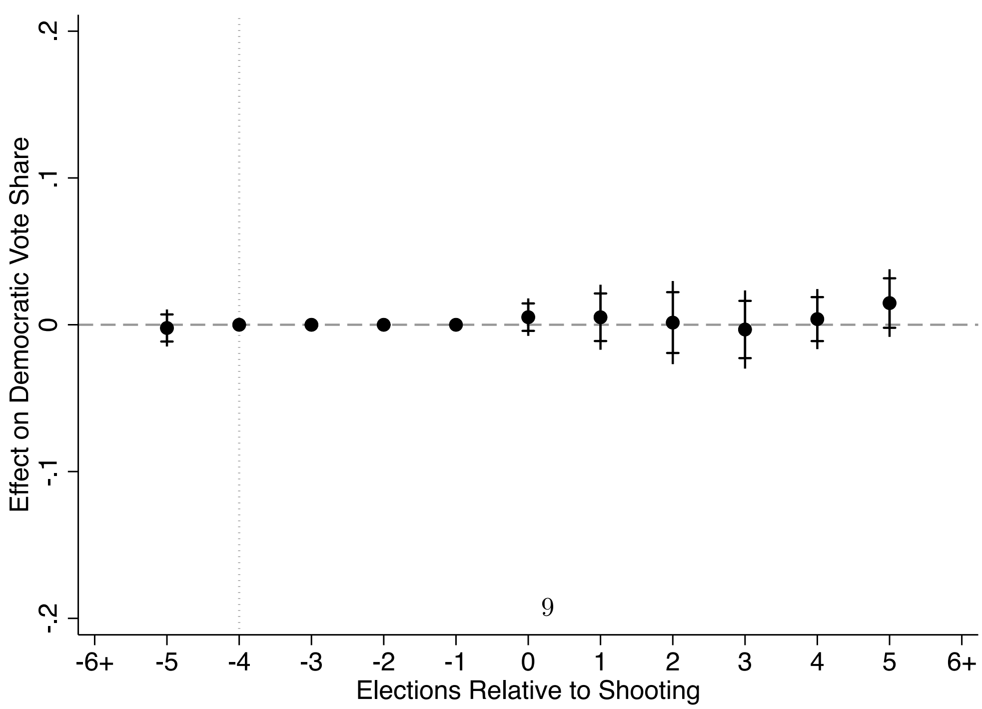
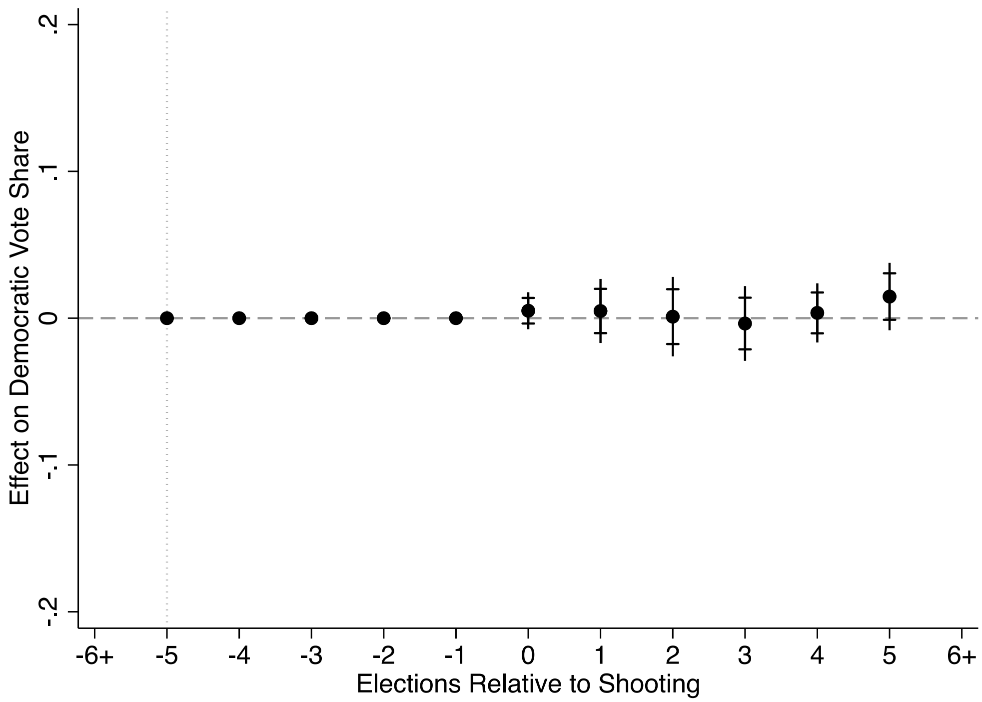


Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021).**Takeaway:** Benchmarked to pre-treatment trends at t-4, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

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

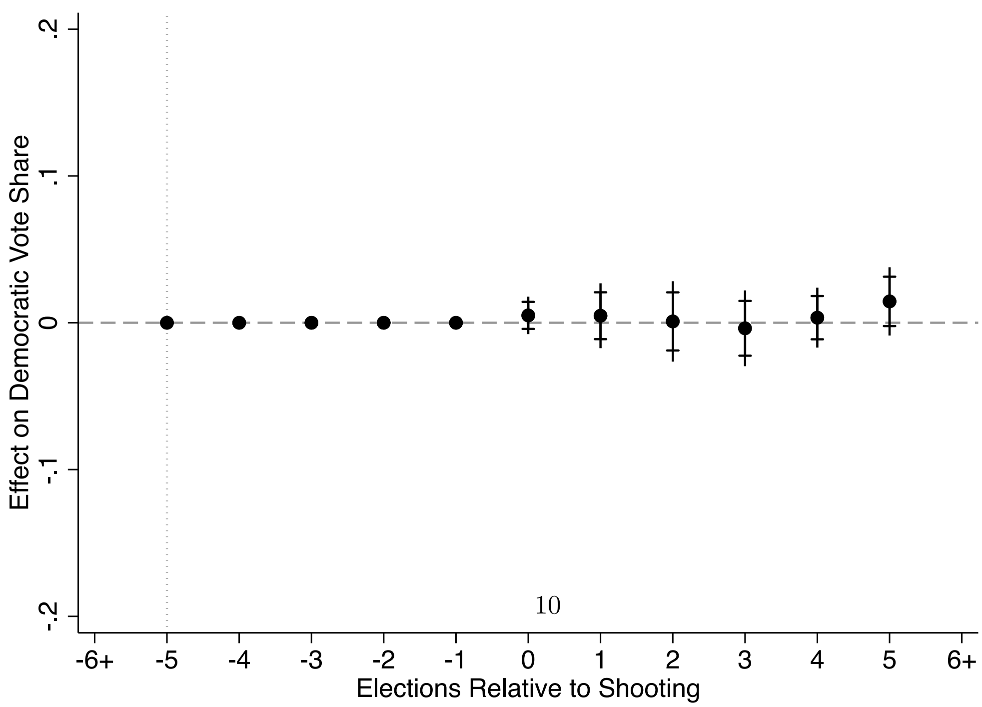


Figure shows the results from using other pre-treatment periods as the baseline as suggested by Freyaldenhoven et al. (2021).**Takeaway:** Benchmarked to pre-treatment trends at t-5, the estimates are even smaller, and even less suggestive of mass shootings having an effect on electoral outcomes.

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

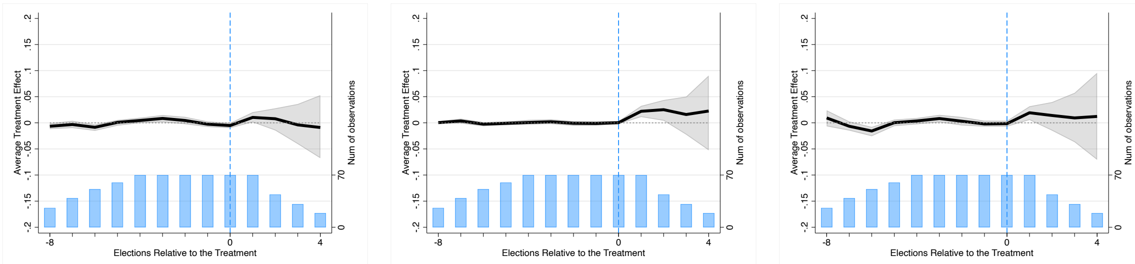


Figure shows the results from using Interactive Fixed Effects Counterfactual Estimator developed by Liu et al. (2021) with different values of r—the number of factors used in estimation—and the integer specifying the order of the polynomial trend term. **Takeaway:** In the interactive fixed effects models, 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

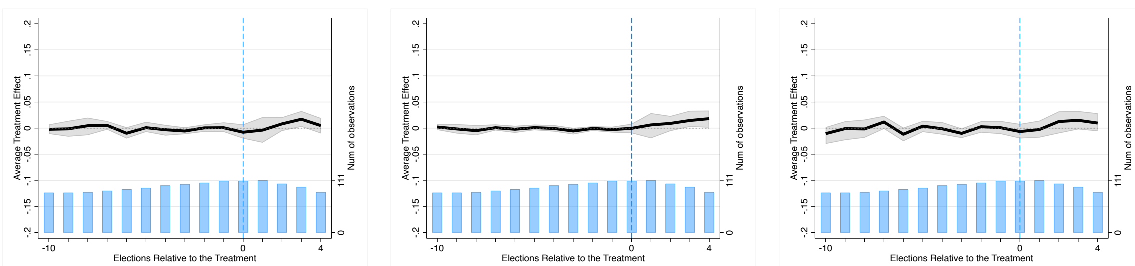
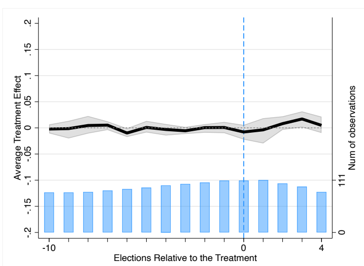


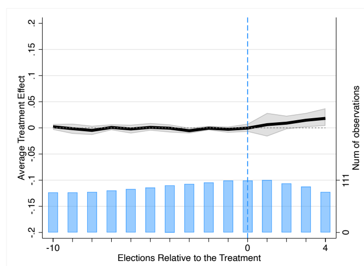
Figure shows the results from using Interactive Fixed Effects Counterfactual Estimator developed by Liu et al. (2021) with different values of r—the number of factors used in estimation—and the integer specifying the order of the polynomial trend term. **Takeaway:** In the interactive fixed effects models, 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

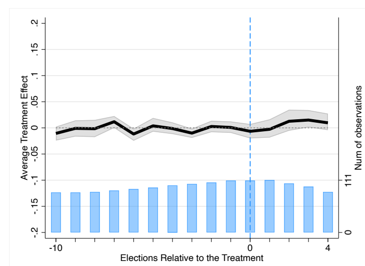
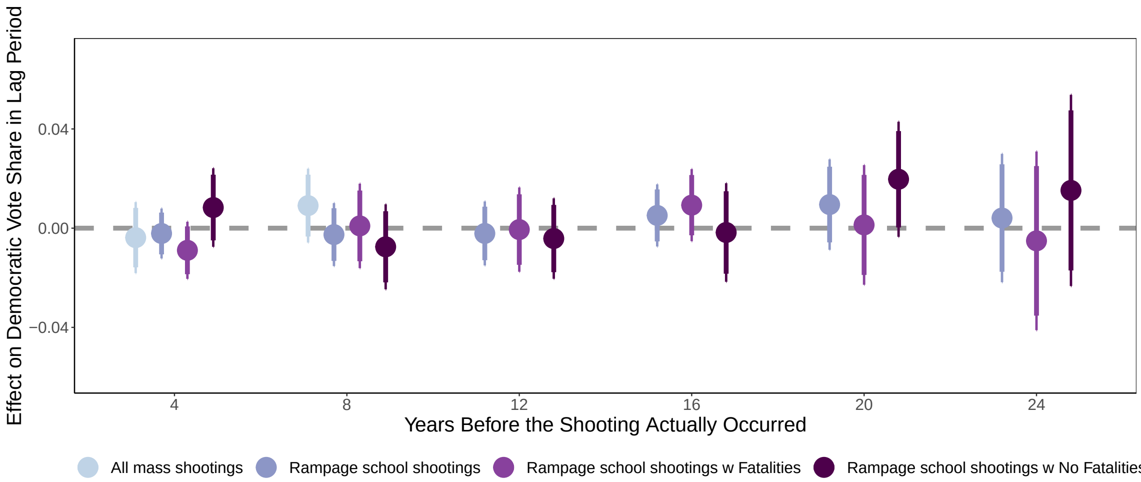


Figure shows the results from using Interactive Fixed Effects Counterfactual Estimator developed by Liu et al. (2021) with different values of r—the number of factors used in estimation—and the integer specifying the order of the polynomial trend term. **Takeaway:** In the interactive fixed effects models, 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

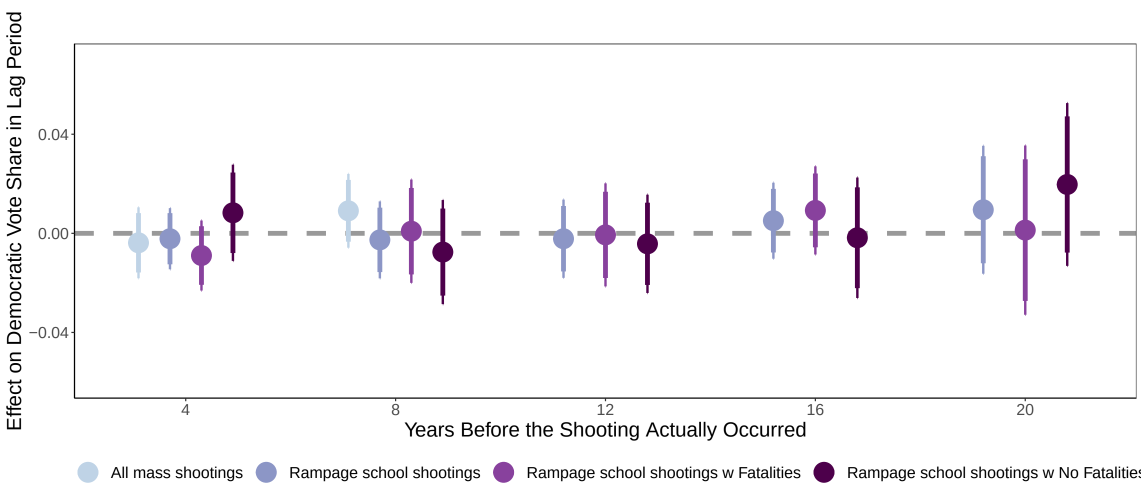
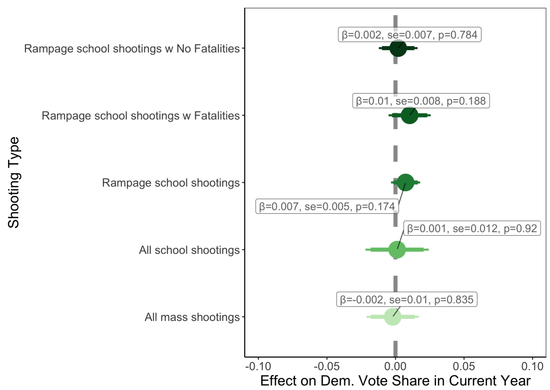


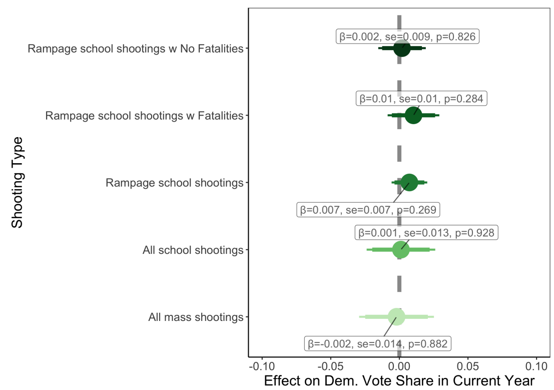
Figure shows the results from using higher order polynomial functional forms for the county-specific trends. The cubic trends model omits the 28 year lag because there are not enough observations in the GMAL data to estimate a model with this many high dimensional fixed effects. The quartic trends model 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

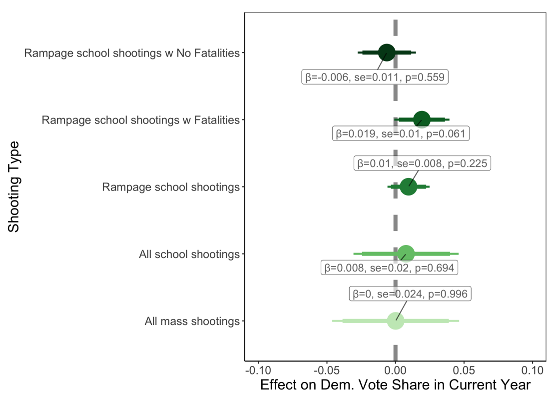
(a) Cubic County Trends Added



(b) Quartic County Trends Added



(c) Cubic County Trends Added, Change in DV



(d) Quartic County Trends Added, Change in DV

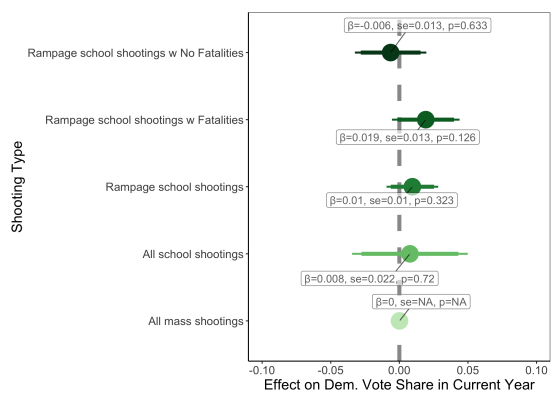
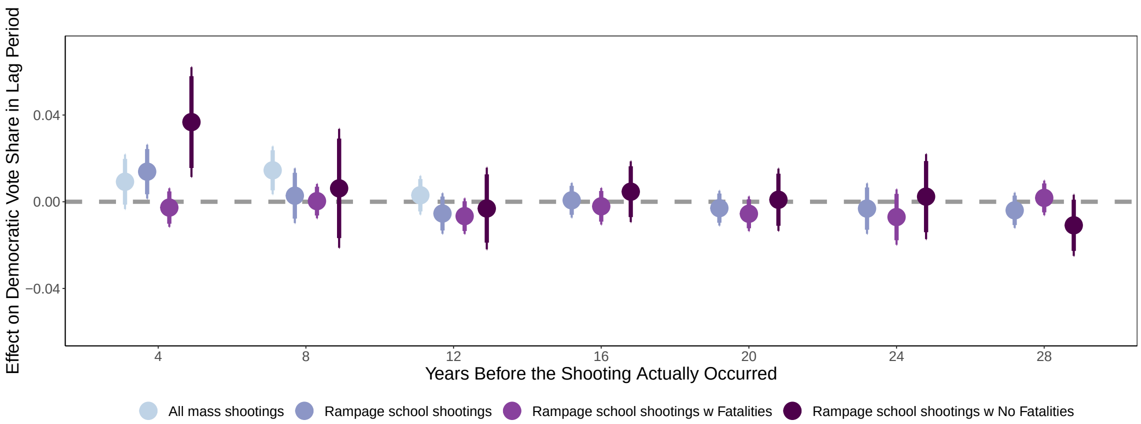


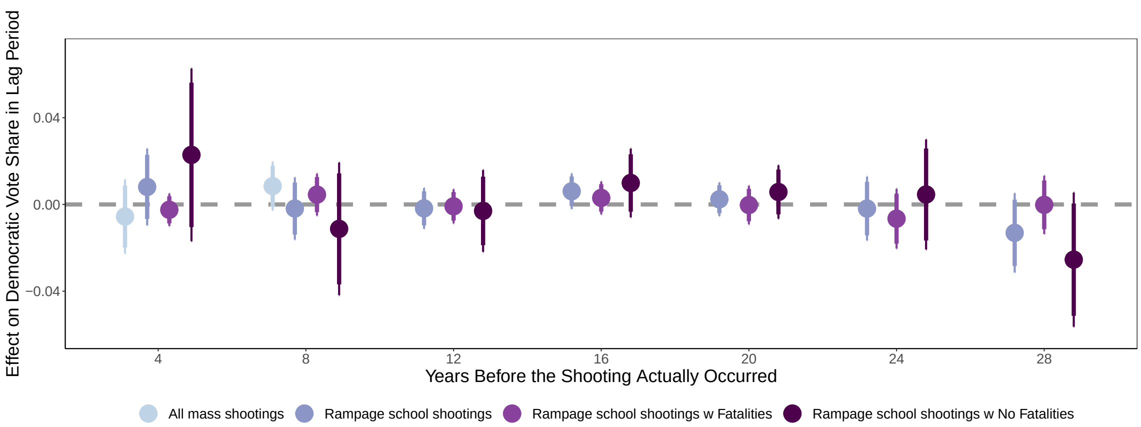
Figure shows the effect of mass shootings of various types once we account for differential trends in Democratic vote share across counties in the United States—this time with cubic and quartic county-specific trends. Within each panel, the first 3 estimates are using the GMAL coding of mass shootings 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

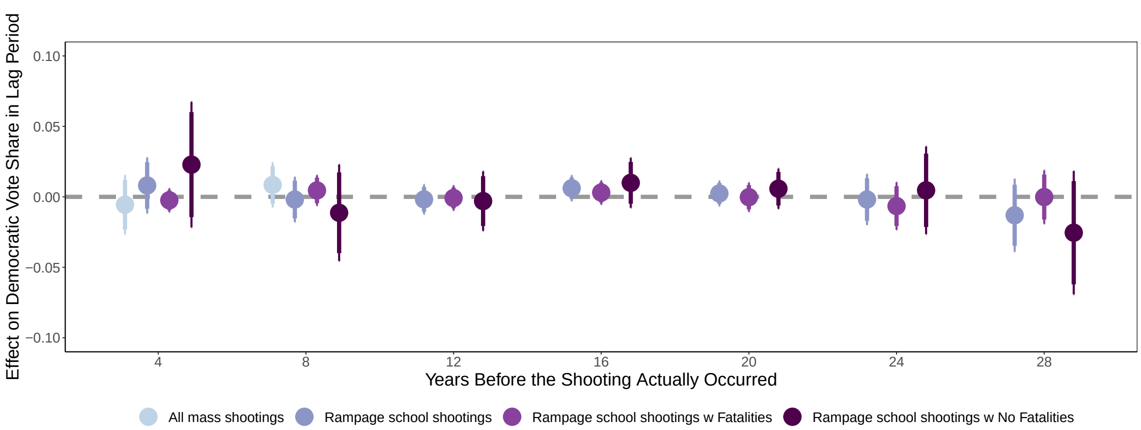
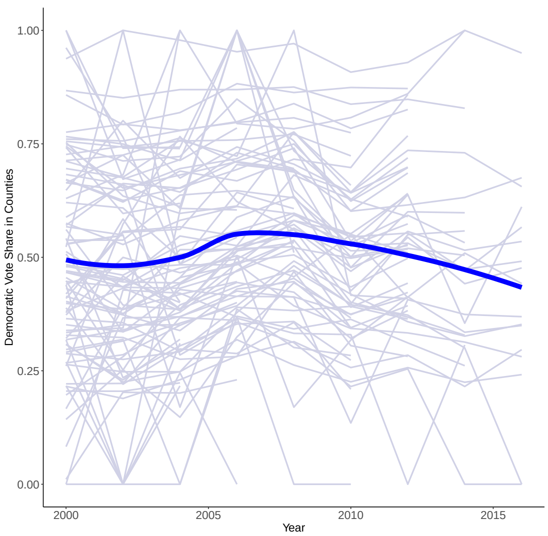


Figure shows the effect of mass shootings on voter turnout in the years prior to when a shooting occurred. **Takeaway:** In contrast to the effects of mass shootings on Democratic vote share which is plagued by trend differences pre-treatment, turnout does not appear to suffer from the same problem,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

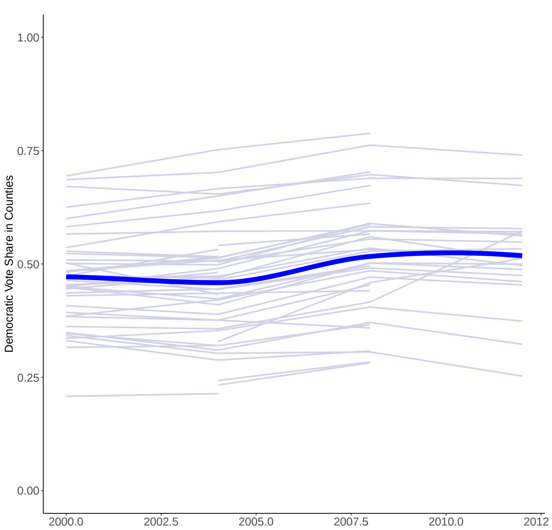
Figure S11: Trends in Presidential Vote Share in Counties With Mass Shootings Prior to These Shootings Occurring, Compared to Trends in Counties Without a Shooting (YOUSAF AND HHB DATA)

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

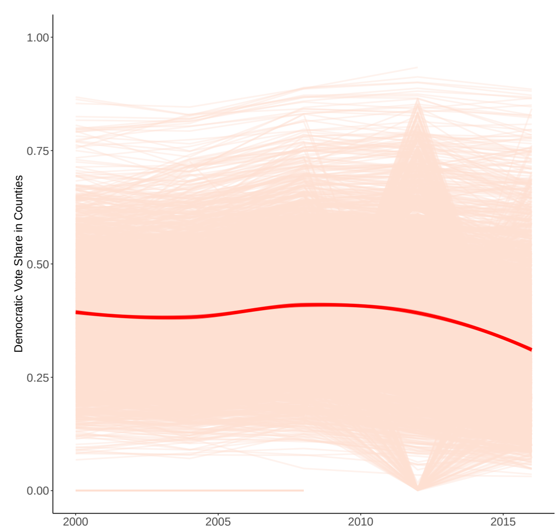


(b) Trends in Democratic vote share in Non Shooting Counties (HHB)

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



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



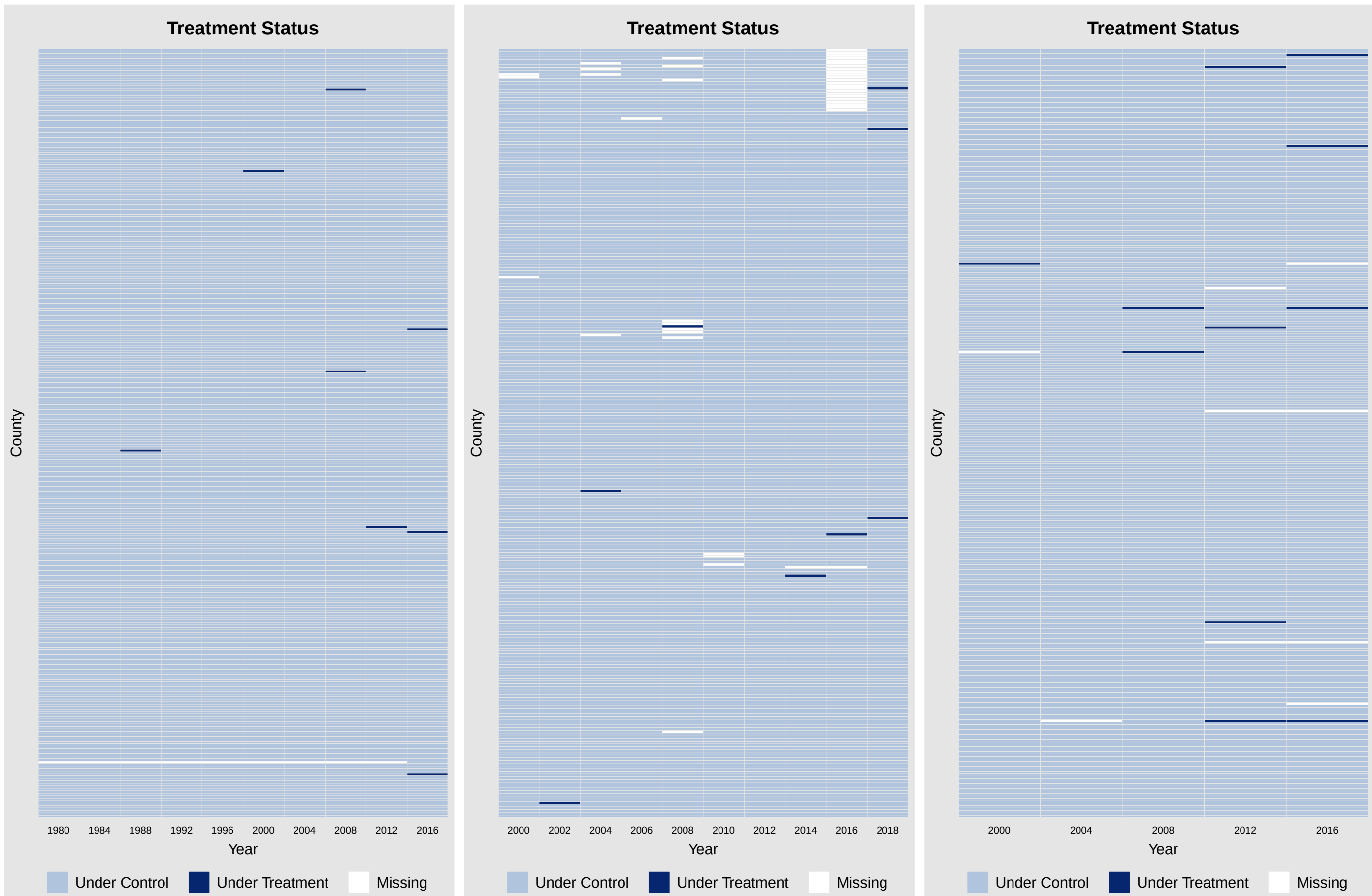
Pre-treatment trends of Democratic vote share in counties where a shooting occurred and benchmarks this to the trends in Democratic vote share found in counties where a shooting did not occur for the Yousaf and HHB data. In the panels on the left, the small blue lines mark the patterns for all counties 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



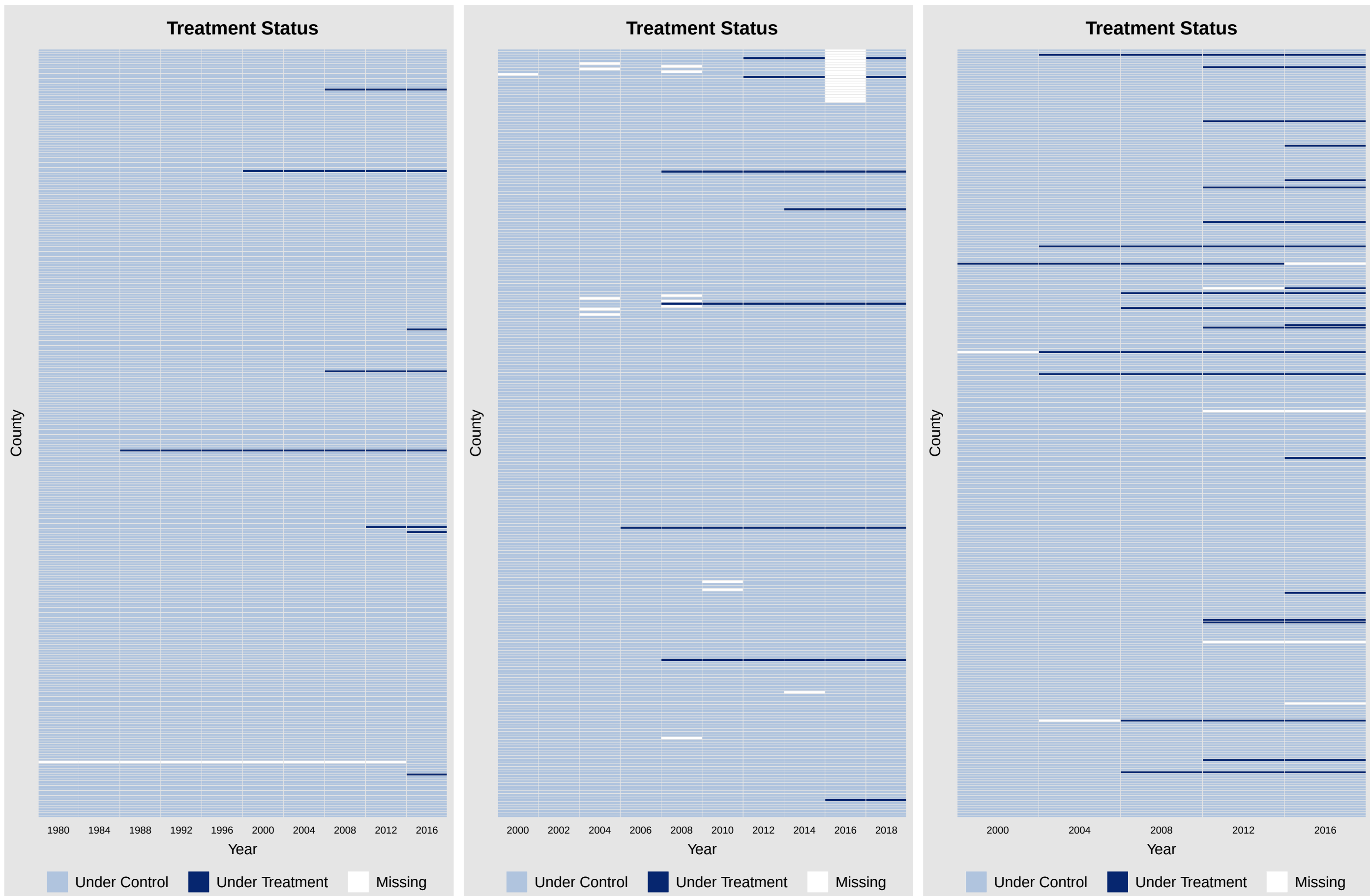
Treatment over time for a random sample of counties in the three datasets illustrating treatment approach 1. Separate random counties are used in the figure that follows this one.

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

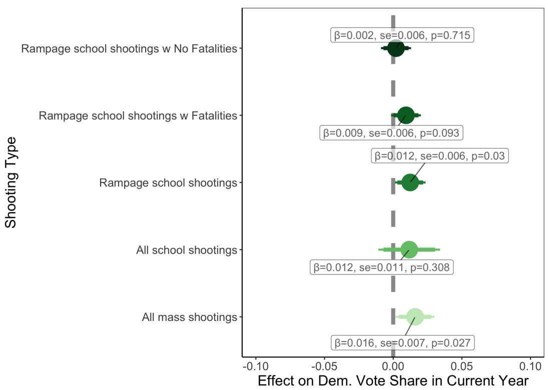
(c) Yousaf Treatment Panel for Random Sample



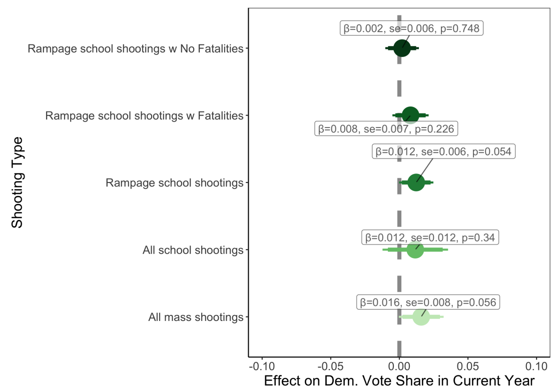
Treatment over time for a random sample of counties in the three datasets illustrating treatment approach 2. Separate random counties are used in the figure that precedes this one.

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

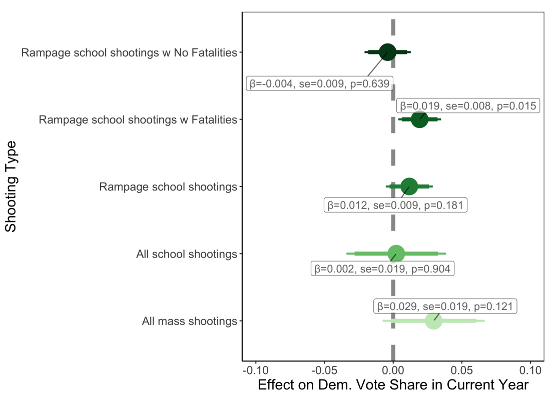


(b) Quadratic County Trends Added

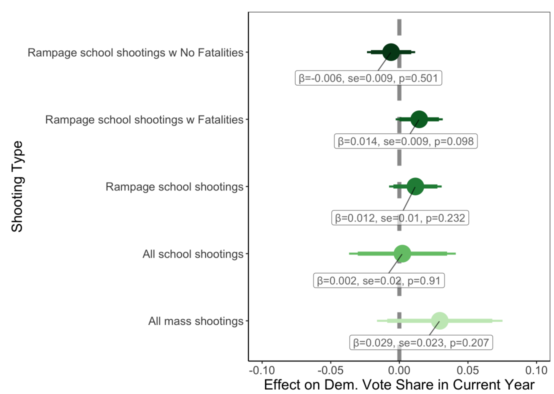


sd

(c) Linear County Trends Added, Change in DV



(d) Quadratic County Trends Added, Change in DV



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

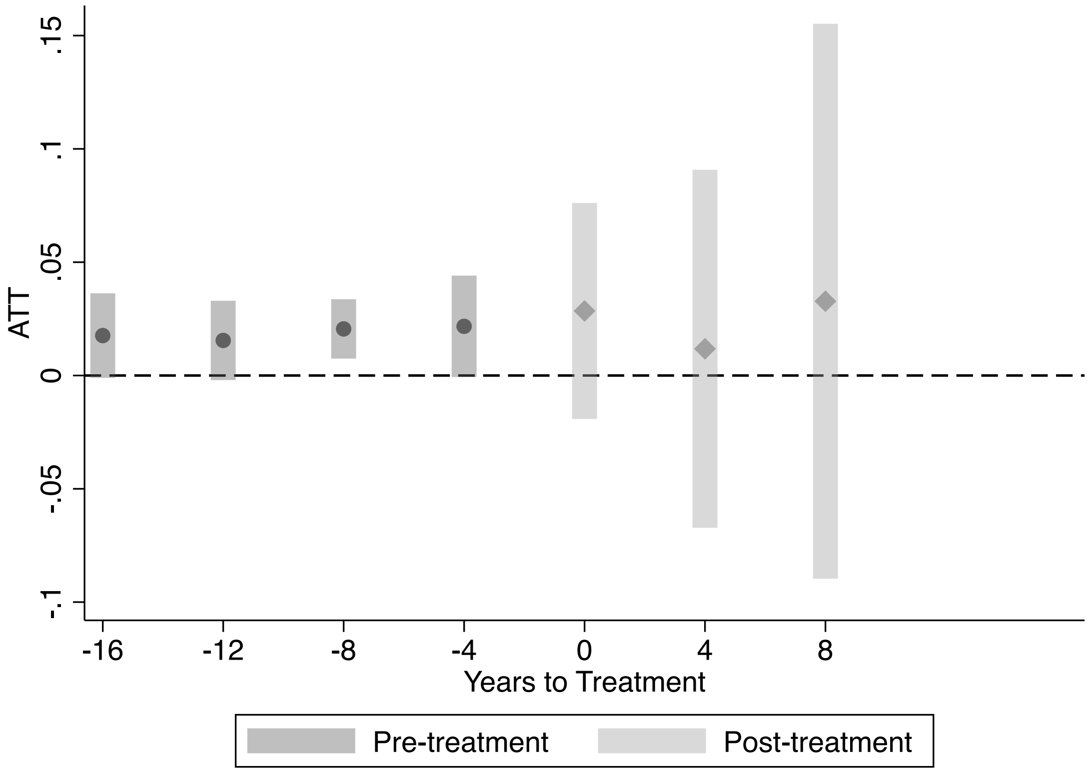
Estimates of the ATT for each period, across all groups or cohorts (i.e. the “Calendar” estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant’Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated 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

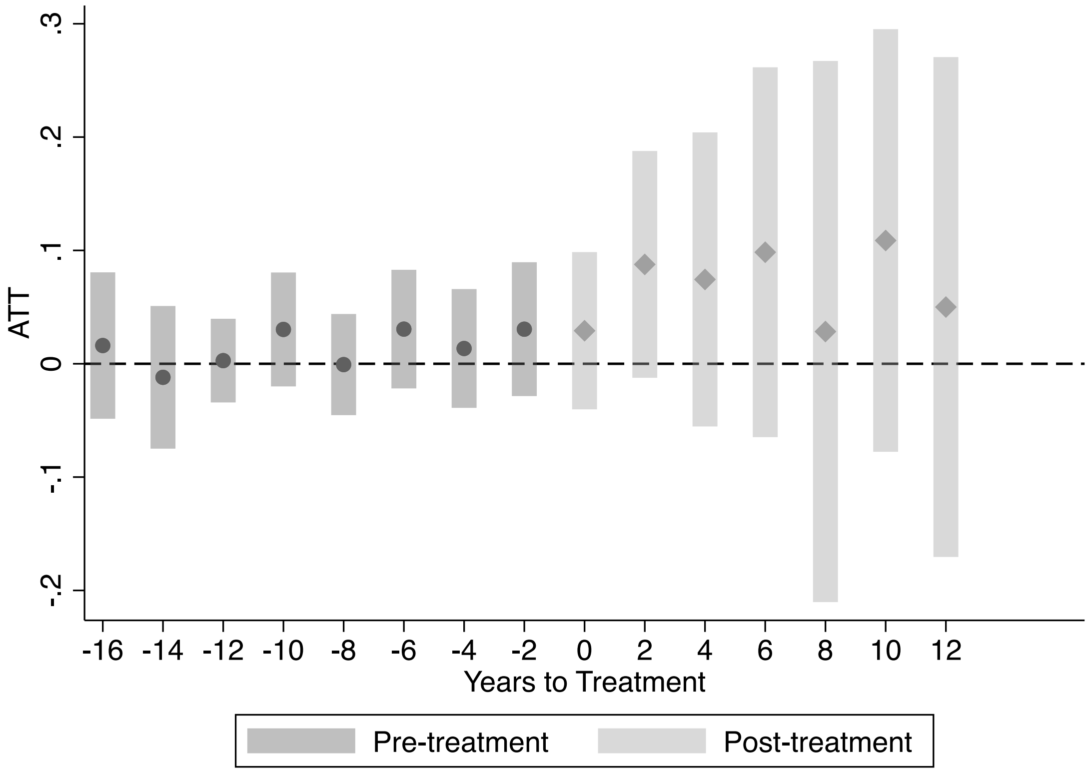
Estimates of the ATT for each period, across all groups or cohorts (i.e. the "Calendar" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated 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

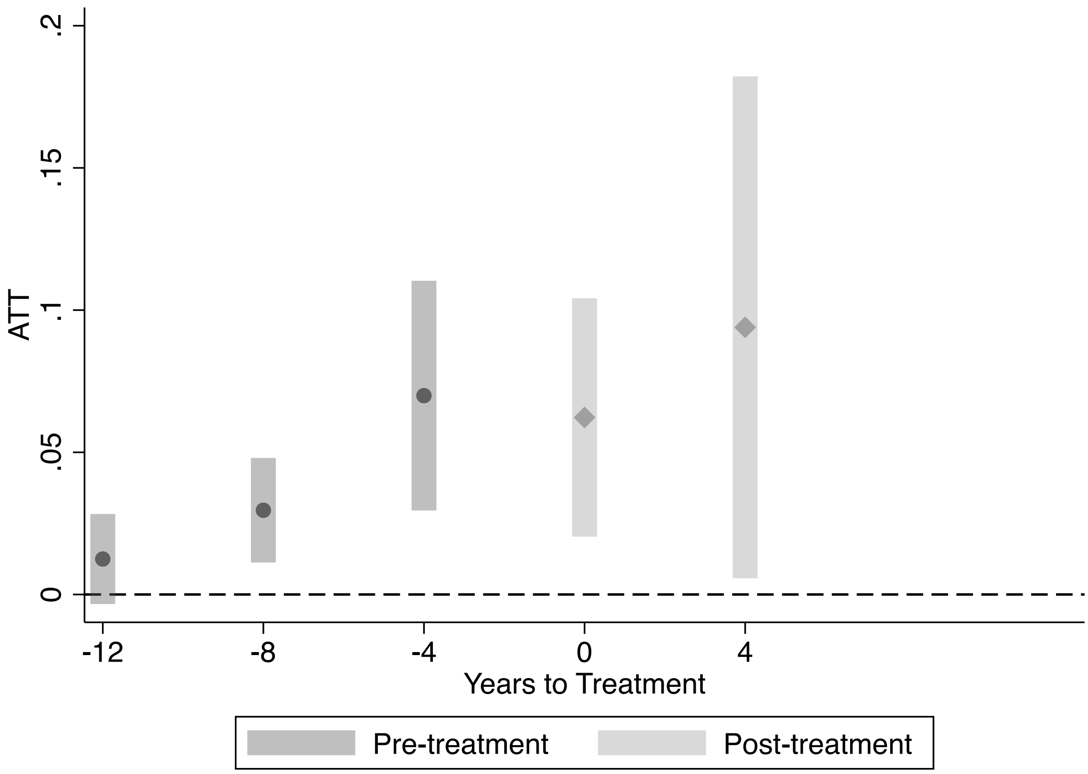
Estimates of the ATT for each period, across all groups or cohorts (i.e. the “Calendar” estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant’Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated 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.

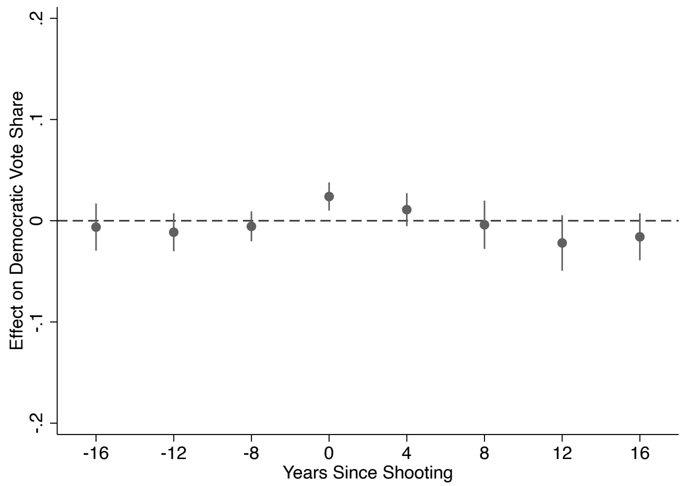
Figure S17: Estimation of all Dynamic Effects (Yousaf)



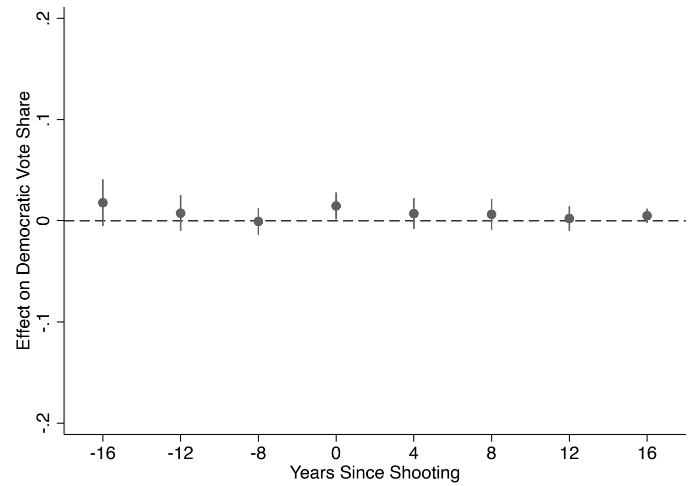
Estimates of the dynamic effects (i.e. the "Event" estimates provided in the CSdid package) based on the procedure developed by Callaway and Sant'Anna (2021). Estimates use the doubly Robust IPW (DRIPW) estimation method, with Wildbootstrap SE, and not-yet treated observations as controls. **Takeaway:** 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)

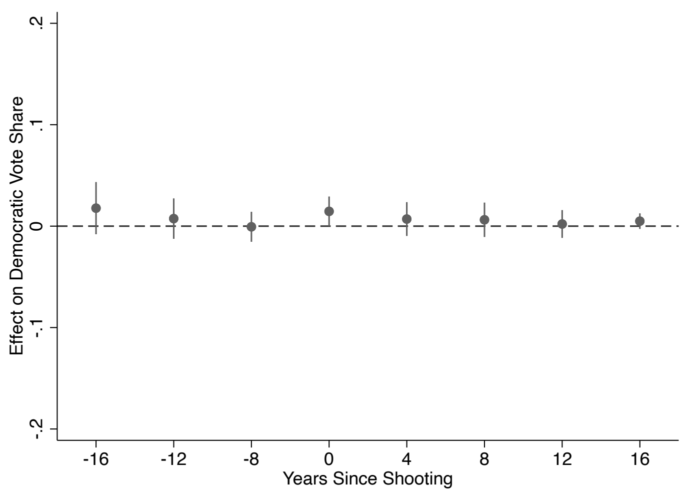
(a) TWFE



(b) Linear Trends



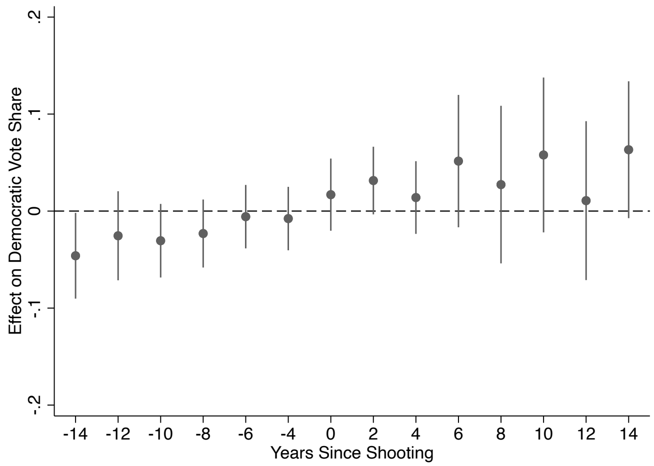
(c) Quadratic Trends



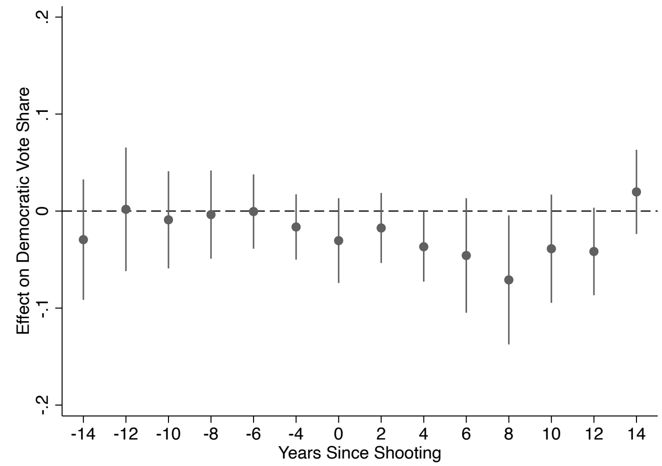
Sun and Abraham (2020) event study estimates through the `eventstudyinteract` package provided by the authors. Standard errors are clustered at the county level. **Takeaway:** Clean comparison effects with trends show no sign of a sizable and durable effect on Democratic vote shares shown in the TWFE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

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

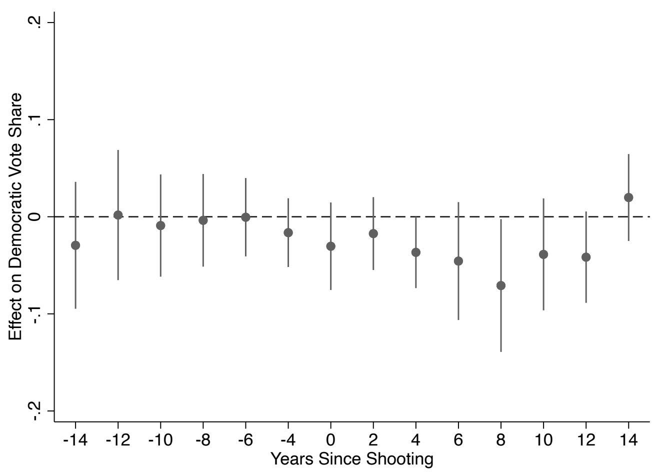
(a) TWFE



(b) Linear Trends



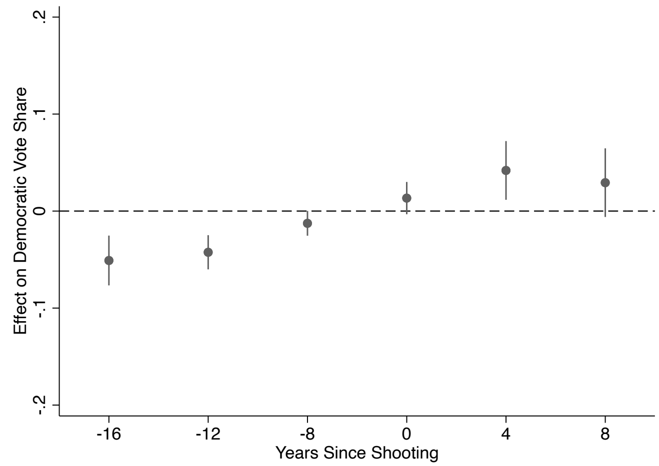
(c) Quadratic Trends



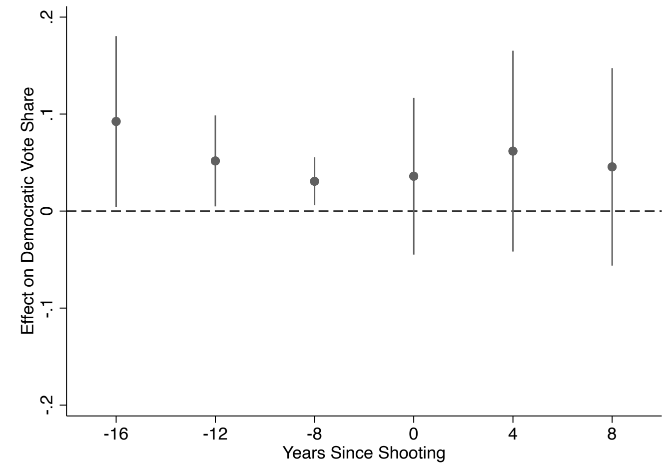
Sun and Abraham (2020) event study estimates through the `eventstudyinteract` package provided by the authors. Standard errors are clustered at the county level. **Takeaway:** Clean comparison effects with trends show no sign of a sizable and durable effect on Democratic vote shares shown in the TWFE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

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

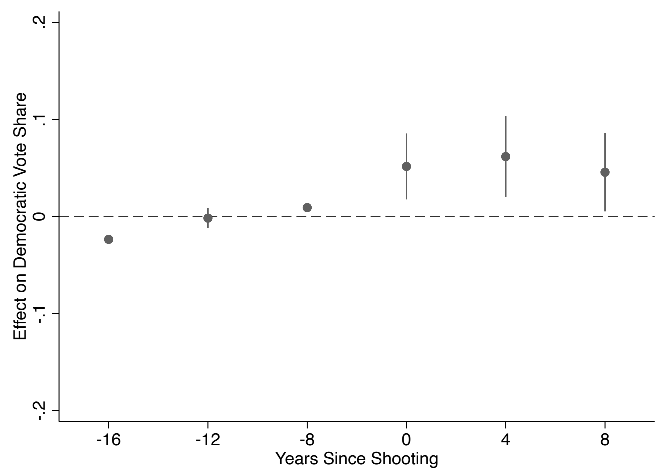
(a) TWFE



(b) Linear Trends

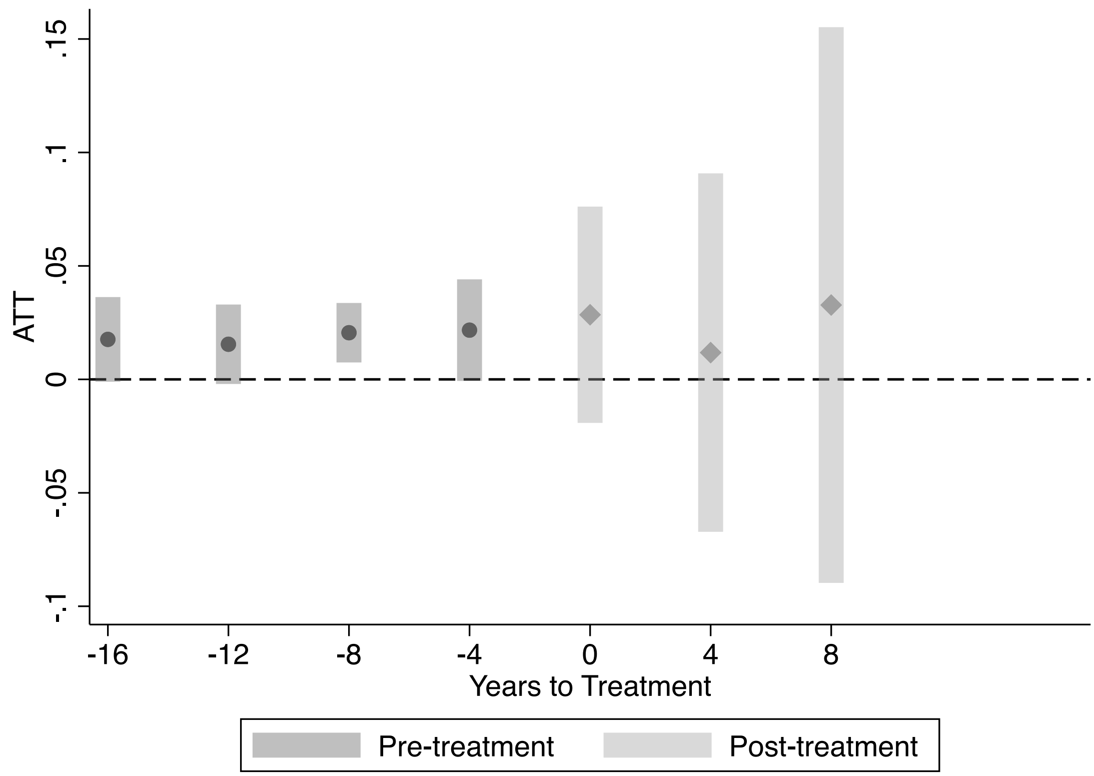


(c) Quadratic Trends



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).

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 S10: 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

Table S11: Adding Year Trends to Covariates in GMAL Data

model	coef	stderr	tstat	pval	N	r2
quad county trends with covs linear trends	.002	.005	.53	.596	18620	.968
linear county trends with covs linear trends	.003	.004	.599	.549	18620	.968
quad trends with covs controlled	.007	.005	1.504	.133	18620	.968
linear trends with covs controlled	.007	.004	1.684	.092	18620	.968
quad trends with no covs omit missing covs	.007	.005	1.578	.115	18620	.967