Do Campaign Visits Increase Local Media Coverage? Evidence from the 2000 Election

Data Task — Predoctoral Application

Corey Gelb-Bicknell

Submitted: Nov 2, 2025

1 Introduction

In this data task, I investigate the causal effect of a candidate's visit to a county on the change in local media coverage that the candidate receives in subsequent days. Intuitively, county visits from political candidates are visible and newsworthy events that could drive an increase in media coverage. Yet, parsing the causal effect of visits on coverage is a non-trivial task. Media coverage is affected by many things unrelated to the visit, and visit locations and timing are chosen in non-random ways that might relate to media coverage.

A naïve approach to this question would compare the average coverage between counties that received a candidate visit and those that did not. This comparison would be very likely to conflate correlation with causation. Among other issues, candidates are likely to visit politically important or media-dense regions like Washington, D.C. As a result, simple cross-sectional comparisons would reflect underlying differences between counties.

Another naïve approach examines average coverage leading up to and after a visit within treated counties. This identification strategy relies on the incredibly restrictive assumption that visit dates are chosen at random. If campaign visits increase in frequency during national political events such as the Democratic National Convention, then the estimated effect of a visit would be confounded by time-varying shocks correlated with visit timing.

To address these concerns and reduce the stringency of our causal assumptions, I employ a TWFE regression to estimate the causal effect with staggered treatment timing, which accounts for differences in media coverage over time and across counties. I implement this estimator separately for each candidate, using a dynamic model with leads and lags to parse out short-run changes in local coverage, as well as a joint model that defines each (county $\times$ candidate $\times$ date) as an observation. It should be noted here that recent advances in econometric research have shown that TWFE suffers from bias under treatment effect heterogeneity. This issue is discussed in detail under the *Methodology* section.

This analysis does not find a statistically significant effect of candidate visits on news media coverage. Non-significant results do not mean that we can accept the null hypothesis, nor do they mean that we should continue to adjust our identification strategy until the results become significant. That said, these null findings should be interpreted with caution. Joint F-tests of pre-treatment coefficients and event-study plots both indicate that treated and control counties were not evolving in parallel prior to the candidate visit, challenging the validity of any causal interpretation. The relatively small sample size of treated counties and the irregular pattern of the coefficients suggest this may come down to issues in our identification strategy or limitations from the small sample.

In the following two sections of this submission, I provide an overview of the methodology utilized in this analysis and review the assumptions required for statistical validity. In the spirit of this data task, I intentionally limit the number of robustness checks employed. Future analyses could extend this work by utilizing a more modern DiD estimator that accounts for treatment effect heterogeneity, re-specifying the regression to measure the percentage change in coverage, assessing heterogeneity in the treatment effect, examining how subsequent visits to a county influence coverage after the initial visit, and explicitly modeling the interaction effects between a candidate and cross-candidate media coverage within the county.

In total, this data task took 25 hours to complete, including time spent reading research papers.

2 Data

This analysis relies on data provided by Jesse Shapiro, which document campaign visits and newspaper article mentions across presidential and vice-presidential candidates during the 2000 election. Each observation represents a (county $\times$ date) pair, with separate columns for visits by each candidate and the frequency of news article mentions of each candidate at both the headline and full-text levels. The raw dataset contains 88 visit-days across 91 unique counties in 22 states. There are 23 unique counties receiving at least one visit and a total of 33,306 (county $\times$ date) observations. A summary of the data's geographic coverage and completeness can be found in Table 2 of the Appendix.

To support my DiD empirical design, I pivot the data long by candidate. Several data adjustments are implemented in order to strengthen causal validity. First, I treat multi-day visits to the same county as a single visit event. For example, former President Bush is recorded visiting Dade, Florida, on Sept. 23 and Sept. 24, 2000. I consolidate these events into one treatment event starting on September 23 and treat it equivalently to a one-day visit. Second, I drop county observations when multiple treatment windows overlap within the same county in order to avoid bias from overlapping treatment exposure. Third, I consider Washington, D.C., to be an outlier county and exclude it from the sample. D.C. received disproportionately more visits than any other county in the dataset, and the nature and motivation of each visit are likely to differ. As the nation's capital, candidates often travel to D.C. for official or logistical reasons that would not constitute a campaign "visit." Furthermore, the D.C. media landscape differs markedly from most local news outlets, since it inherently focuses on national political coverage. A summary table of the pre- and post-processing visit and media coverage data is presented in Table 3 of the Appendix.

3 Methodology

For this analysis, I use a staggered, dynamic TWFE model with county and day fixed effects to measure the causal effect of a candidate visit on the total count of newspaper mentions of that candidate. This estimate is calculated separately for each candidate.

A staggered framework extends the classic DiD approach to a setting where treatment timing varies across units in the sample. In our context, the treatment occurs when a candidate visits a county, with counties treated at different points over the course of the campaign. County fixed effects account for time-invariant differences in baseline coverage, and day fixed effects account for national-level changes in media coverage over time.

Our model is dynamic, meaning that it has leads and lags to trace the evolution of the effect before and after a visit. This allows us to observe potential violations of our causal assumptions (such as no anticipation and parallel trends) and observe how quickly the effect fades afterward. Because we expect that visits only have a short-term effect on coverage, I restrict the estimation window to five days before and after each visit. Critically, I assume that the treatment effect is temporary and that treated counties return to an untreated state within five days of the visit. Subsequent visits to the same county are treated equivalently as additional treatments to the county.

Following the guidance of Abadie et al. (2017), I cluster standard errors at the county level. Clustering is only appropriate when sampling or treatment assignment occurs at the cluster level, since observations within clusters are not statistically independent. In our case, the treatment (candidate visit) is applied at the county level, so each observation is correlated at the county level. Failing to cluster would underestimate the true uncertainty of our estimate, since the variance within each county is artificially low.

Formally, our specification for each candidate c is given by:

$$Y_{i,t,c} = \alpha_i + \lambda_t + \sum_{k=-5}^{5} \beta_k D_{i,t+k,c} + \varepsilon_{i,t,c},$$

where $Y_{i,t,c}$ denotes total newspaper mentions for candidate c in county i on day t. α_i and λ_t are county and day fixed effects, and $D_{i,t+k,c}$ is an indicator equal to one if county i is k days away from its first visit by candidate c. The coefficients β_k represent the dynamic treatment effect relative to the visit date, where $(k < 0)$ captures pre-trends and $(k > 0)$ represents post-treatment effects.

As an extension of this analysis, I also complete a pooled specification where each county-candidate-day is treated as a unique observational unit. By including all candidates within the same specification, we increase the sample size of the treated group and the statistical power of our estimate at the cost of more restrictive causal assumptions. Specifically, this pooled model requires that candidate coverage and visits are assumed to be independent, such that a visit by one candidate does not induce coverage of another. I add fixed effects for candidate and cluster standard errors by (county $\times$ candidate), since treatment is assigned at the county-candidate level. I am less confident in the validity of this analysis.

In recent years, traditional staggered TWFE DiD designs have fallen out of favor because of research showing they produce biased estimates under heterogeneous effects. As Goodman-Bacon (2019) demonstrates, the TWFE estimator combines a series of two-by-two DiD comparisons, some of which use earlier-treated units as controls for later-treated ones. These comparisons are valid only when the treatment effect is homogeneous (i.e., identical across cohorts and over time). When the treatment effect varies, these comparisons mix units that are no longer valid counterfactuals, and the resulting weights can even be negative. This can bias, attenuate, or sign-reverse the estimate.

Despite these theoretical concerns, I utilize the TWFE DiD design here in the spirit of the data task. This estimator is potentially less problematic in this setting because the treatment effect of a candidate visit on newspaper coverage is expected to be transient over a short time horizon. Assuming that previously visited counties return to an untreated state, the vast majority of control observations for each cohort are genuinely untreated during the event horizon window, mitigating the bias described above. Future work should repeat this analysis under a specification that explicitly allows for heterogeneity in the treatment effect.

One additional limitation of this analysis is that the outcome variable is measured in raw media coverage counts rather than relative changes in coverage over time. Because baseline coverage varies significantly across counties, the relative change in coverage may be more meaningful than the total change. An improved analysis would transform the outcome variable (coverage) using a Poisson regression model. The coefficients should be interpreted as the relative percentage change in coverage. Per Chen and Roth (2024), log-linear transformations should not be used due to the significant presence of observations with zero coverage, which forces us to drop many observations or distorts the scale and biases our estimates when we add one.

4 Assumptions

Our identification strategy relies on several key assumptions for the coefficient to truly capture the causal effect of a candidate visit on local news coverage. Below, I outline the assumptions and describe robustness checks designed to assess validity.

4.1 Parallel trends

The parallel trends assumption requires that, absent treatment, both treated and untreated counties would follow similar trajectories in news coverage over time. Formally, for each unit that receives treatment,

$$\mathbb{E}[Y_{i,t}(0) - Y_{i,t-1}(0) \mid D_{i,t} = 1] = \mathbb{E}[Y_{i,t}(0) - Y_{i,t-1}(0) \mid D_{i,t} = 0]\,,$$

for all t, where $Y_{i,t}(0)$ represents the potential outcome for county i at time t in the absence of a candidate visit, and $D_{i,t}$ is an indicator for the treatment (equal to one only if a candidate visited county i by time t).

This assumption would be violated if news coverage evolved differently across counties. An obvious concern arises when multiple candidates visit the same county within the same treatment window. If visits by one candidate affect the coverage of another, then there is an exogenous confounder influencing the outcome variable independent of our treatment. Because I estimate the effects of each candidate separately, these cases represent a violation of parallel trends rather than spillover of the treatment. An extension of this analysis could control for visits from other candidates.

A common, traditional diagnostic to validate the parallel trends assumption is to examine the pre-treatment dynamics using the leading coefficients prior to event timing. If parallel trends hold, we would expect the pre-treatment coefficients to be statistically indistinguishable from 0, both at the individual level as well as through a joint test. Implicit in this diagnostic is that there are no time-dependent confounders that impact outcomes in the post-treatment period. If both candidate visits and news coverage are driven by exogenous political events that drive coverage regardless of a candidate's presence, then parallel trends would be violated even if pre-trends held.

Moreover, Roth (2022) and others have shown that pre-trend testing suffers from low statistical power. Furthermore, Roth has found that conditioning the analysis on the result of a pre-test can bias estimation, since it serves as a form of sample selection on a random event (samples that pass pre-tests are more likely to have bias in the pre-treatment coefficients that also applies to post-treatment coefficients). Furthermore, it is implausible to assume perfect parallelism in practice. It's very unlikely that media coverage evolves the same across counties over time.

A more transparent approach to testing parallel trends is to quantify the robustness of the results conditional on deviations from parallel trends. Rambachan & Roth (2023) have developed an Honest DiD framework that allows us to assess how significant a violation the parallel trends assumption would need to be for the causal estimate to lose statistical significance. With more time, I would calculate and report these numbers.

With that said, traditional methods for assessing pre-trends suggest serious violations of the parallel trends assumption. As shown in the results section, pre-treatment coefficients vary wildly, and joint F-tests fail in our smaller samples.

4.2 No anticipation

A second core identifying assumption in the DiD framework is the requirement for no anticipation of the treatment. We would expect that treatment in a given period does not influence the outcomes in prior periods. Formally,

$$Y_{i,t} = Y_{i,t}(0) \quad \text{for all } t < T_i,$$

where T_i denotes the first day county i receives a candidate visit, and $Y_{i,t}(0)$ denotes the potential outcome in the absence of treatment.

In our context, the no-anticipation assumption warrants particular attention. Most candidate visits are likely scheduled far in advance, and we would expect local newspapers to be aware of upcoming events. As a result, media outlets may begin to increase coverage prior to the visit day in anticipation of that event. For example, outlets may cover community preparations, protests, or official statements from the local government in the days leading up to a candidate's arrival.

If anticipation effects exist within the treatment window, it would manifest as deviations in the pre-treatment coefficients of our specification. Although parallel trends seem to be violated, there is no obvious evidence of significant pre-treatment effects. For further robustness, one could extend the pre-treatment window and shift the treatment onset earlier to account for some level of anticipated coverage.

4.3 No spillovers

The third core identifying assumption in our DiD framework is the absence of spillover effects. It requires that treatment of one unit does not affect the potential outcomes of another. Formally, for all $i \neq j$,

$$Y_{it}(d_i, d_j) = Y_{it}(d_i),$$

where $Y_{it}(d_i, d_j)$ denotes the potential outcome for unit i when unit j receives treatment d_j.

In this setting, spillover effects are a significant concern. A candidate visit may receive coverage in adjacent counties. Worse, a high-profile visit event may receive national news, with the effect being absorbed by our day fixed effect, contaminating the control group and biasing our estimate downward.

One way to empirically assess the validity of this assumption is by checking whether treatment in one county affects news coverage in surrounding counties. If we could accurately quantify the degree of treatment exposure, the binary treatment variable could be replaced with a continuous measure of treatment intensity to account for partial spillovers. To conduct a rudimentary robustness test that mitigates the concern of spillovers, I re-estimate our model excluding county days within the same state during a treatment window, reducing the likelihood of contamination of the treatment in neighboring counties. The results are excluded from the paper for brevity but produce similar results.

4.4 Minimal Contamination from Treated Controls

As described above, recent literature has found that the TWFE estimator produces biased estimates under treatment heterogeneity. We rely on the assumption that contamination of treated controls is limited in our analysis due to the high volume of untreated states, as well as the assumption that the treatment effect is transient. This assumption is both fragile and difficult to validate. In practice, one could partially assess plausibility by verifying that treatment effects dissipate quickly after treatment.

4.5 No Selection on Treatment Effect Heterogeneity

Even if the DiD causal assumptions described above hold exactly, we are only guaranteed internal validity of our estimate. If candidates exhibit selection bias on treatment effect heterogeneity (i.e., choosing to visit counties where they anticipate a higher return on media coverage), then our effect size will not generalize across counties not visited within our sample. To address this concern, extensions to this analysis could construct a weighted average of the heterogeneous treatment effect based on observable county characteristics to produce a more generalizable estimate of effect size. Alternatively, we could seek to model heterogeneity explicitly, clarifying how the media impact of a candidate visit varies across county characteristics.

5 Results

The results of the staggered dynamic DiD estimates are shown in the plots and table below. The results provide limited evidence of a consistent causal effect of candidate visits on local media coverage. Across all five specifications, the pre-treatment coefficients fluctuate substantially, suggesting that parallel trends may be violated. For Bush and Gore, most of the pre-treatment coefficients are statistically insignificant, and several post-treatment coefficients are positive and moderately large, indicating a possible transient increase in local coverage following a visit. For Cheney, the effects are significant in both the pre- and post-treatment periods, and the joint F-statistic for pre-treatment coefficients is highly significant ($p < 0.01$), suggesting coverage was already trending positively prior to the treatment. Lieberman has consistently negative, statistically significant coefficients for most of the event-study window, indicating systematic differences in coverage rather than causal effects driven by treatment.

The pooled regression model shows a modest positive post-treatment effect, but the effects have wide confidence intervals and are small in magnitude. Pre-treatment trends are jointly non-significant, but multiple pre-treatment coefficients are significant when considered in isolation. Taken together, the results of this analysis indicate no robust or consistent evidence that a candidate visit increases newspaper coverage of that candidate. Instability in the coefficients, as well as significant pre-trends for some candidates, point to potential violations of the key causal assumptions rather than true treatment effects.

Figure 1: Event study coefficient plots by candidate

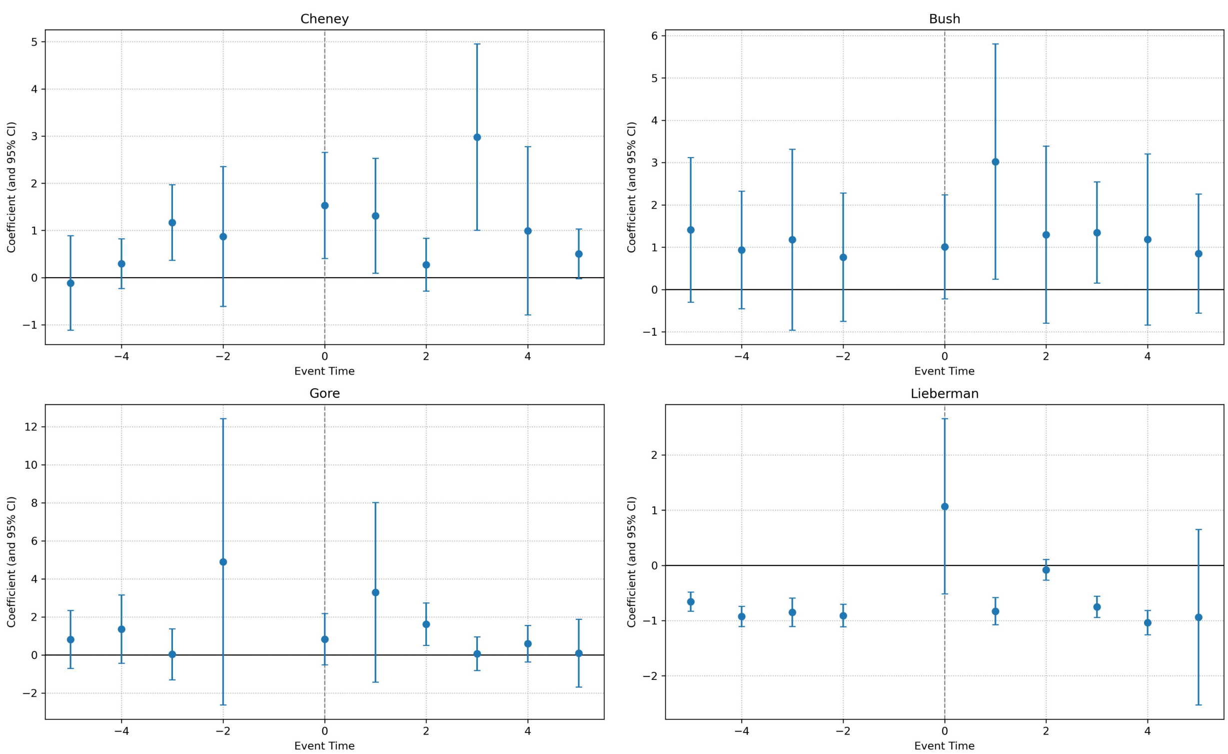

Table 1: Dynamic Event-Study Estimates by Candidate

	Dependent variable: counts				
	Bush	Cheney	Gore	Lieberman	All
	(1)	(2)	(3)	(4)	(5)
Event -5	1.408	-0.115	0.818	-0.659***	0.970*
	(0.872)	(0.510)	(0.778)	(0.089)	(0.550)
Event -4	0.933	0.294	1.352	-0.926***	1.048*
	(0.709)	(0.270)	(0.916)	(0.093)	(0.614)
Event -3	1.176	1.166***	0.033	-0.850***	0.417
	(1.091)	(0.408)	(0.686)	(0.131)	(0.391)
Event -2	0.762	0.869	4.895	-0.910***	2.758
	(0.775)	(0.756)	(3.837)	(0.104)	(1.749)
Event 0	1.008	1.529***	0.834	1.067	1.063**
	(0.627)	(0.573)	(0.688)	(0.809)	(0.495)
Event +1	3.022**	1.308**	3.295	-0.829***	2.865
	(1.418)	(0.622)	(2.409)	(0.125)	(1.829)
Event +2	1.292	0.272	1.612***	-0.081	1.614**
	(1.068)	(0.284)	(0.569)	(0.096)	(0.762)
Event +3	1.346**	2.977***	0.071	-0.750***	0.525
	(0.610)	(1.007)	(0.451)	(0.098)	(0.546)
Event +4	1.182	0.990	0.592	-1.039***	1.177**
	(1.031)	(0.909)	(0.487)	(0.112)	(0.498)
Event +5	0.847	0.504*	0.088	-0.940	0.368
	(0.717)	(0.268)	(0.909)	(0.810)	(0.480)
Date FE	x	x	x	x	x
County FE	x	x	x	x	x
Candidate FE					x
Clustered SEs	County	County	County	County	County × Cand
Pre-trend F-stat (p-val)	1.21 (0.314)	39.58 (0.000)	1.69 (0.158)	53.19 (0.000)	1.12 (0.352)
Observations	33191	33191	33191	33191	132764

Note: *p<0.1; **p<0.05; ***p<0.01
Table reports event-study coefficients from separate regressions by candidate.
Standard errors clustered by county in parentheses

6 Appendix

Table 2: Geographic Coverage and Completeness of Sample

	Value	**Definition**
Total states	22	Number of distinct states in sample
Total counties	91	Number of distinct counties in sample
Average counties per state	4.1	Mean number of counties per state in sample
County-days observed	33,306	Total number of observations (county × date)

Table 3: Major Candidate Visits and Media Coverage, 2000 Election

	Bush	Cheney	Gore	Lieberman	**All**
Full Dataset					
Total Visit Days	37	5	41	5	**88**
# Unique Visits	31	5	39	5	**80**
# Unique Counties Visited	23	5	14	5	**28**
Annual Coverage per County	1,436	164	1,188	116	**2,903**
Annual Coverage per Visited County	1,763	450	1,783	87	–
After Data Cleaning					
Total Visit Days	33	5	24	3	**65**
# Unique Visits	28	5	22	3	**58**
# Unique Counties Visited	22	5	12	3	**27**
Annual Coverage per County	1,436	164	1,188	116	**2,903**
Annual Coverage per Visited County	1,813	450	1,934	84	–