Quarterly Journal of Political Science, 2024, 19: 307–330

Do Billboard Advertisements Increase Voter Turnout? A Large-Scale Field Experiment

Donald P. Green[1], Lionel Ong[1], Kylan Rutherford[1] and Aaron Schein[2*]

[1] *Department of Political Science, Columbia University, New York, NY 10027, USA; dpg2110@columbia.edu*

[2] *Department of Statistics, The University of Chicago, Chicago, IL 60637, USA; schein@uchicago.edu*

ABSTRACT

Although an extensive experimental literature has tested a wide array of voter mobilization tactics, billboard advertisements have seldom been evaluated, and studies to date have been limited to a small number of sites. This essay reports results from a nationwide experiment conducted during the 2020 general election in the United States. Experimental sites were ethnically diverse locations in metro areas, including both presidential battlegrounds as well as places with no closely contested races. A total of 298 billboards were randomly assigned to treatment or control in 155 geographic clusters. Exposure to billboards by residential location is modelled using cell phone usage patterns. Turnout is measured using public records for residents living at various distances from randomly assigned billboards. Using a variety of estimation approaches, we obtain point estimates that are close to zero, with hints of stronger effects among those who reside near treated billboards.

*The authors thank Nate Persily and Charles Stewart, who made this project possible, and Greg Huber, who offered helpful design suggestions. We thank Turnout Nation, which paid for the billboards and data collection, and Veraset, a company that aggregates anonymized population movement data.

The authors have no financial or family ties to Veraset, Turnout Nation, or outdoor advertising companies. This project was reviewed and approved by the Institutional Review Board at Columbia University, protocol AAAT-4005.

Online Appendix available from:
http://dx.doi.org/10.1561/100.00022062_app
Supplementary Material available from:
http://dx.doi.org/10.1561/100.00022062_supp
MS submitted on 27 May 2022; final version received 5 June 2023
ISSN 1554-0626; DOI 10.1561/100.00022062
© 2024 D. P. Green *et al.*

On the whole, it appears that signage does little to raise turnout in high-salience elections.

Keywords: Voter turnout; elections; voter mobilization; billboards; advertising

Signage by candidates for public office is a prominent feature of election campaigns in many countries, so much so that some places (e.g., Spain) regulate it to ensure that all parties have equal access to desirable locations (Esteban-Casanelles, 2020). Signage in United States is lightly regulated and attracts relatively few campaign dollars, especially by comparison to other advertising media, such as television, on which campaigns spend lavishly (Guzzetta, 2006, p. 187).

One reason why American campaigns place relatively little emphasis on signage is ambiguity about whether signage is effective. Leading how-to books for campaign managers are generally lukewarm about the merits of buying signage, particularly large-format billboards or roadside signs (Blodgett and Lofy, 2008, p. 133; Shaw, 2014, p. 324). Signage is not dismissed as ineffective, but compared to other tactics, such as online advertising, is discussed with little enthusiasm.

The effects of signage remain ambiguous as we turn our attention from candidate campaigns to nonpartisan get-out-the-vote (GOTV) efforts. From a theoretical standpoint, it is unclear whether billboard messaging should be presumed effective because it conveys a "noticeable reminder" about the upcoming election (Dale and Strauss, 2009) or ineffective because impersonal encouragements to vote (e.g., pre-recorded phone calls) tend to have little effect (Green and Gerber, 2019). The empirical literature also paints a mixed picture. An early billboard experiment conducted in 2007 found municipal turnout in three treated cities to be almost identical to turnout in 21 control locations (Panagopoulos and Ha, 2015). This study was replicated on a larger scale by Minkoff and Mann (2020), who randomized the location of Vote.org billboards within metropolitan areas in four states that held elections in 2019. Billboards were found to increase turnout by an average of 2.3 percentage points over an 18.4% base rate of voting in the control group.[1]

The present study represents the first large-scale evaluation of GOTV billboards conducted in a presidential election. We begin by discussing the competing theoretical perspectives on whether signage would be expected to

[1] By contrast, MoveOn's nonpartisan yard sign campaign, which was designed to encourage voter turnout in the 2018 midterm election, had no effect (Green and Gerber, 2019, chap. 4). Four prior experiments on candidate yard and roadside signs found no turnout effects (Green *et al.*, 2016).

increase voter turnout. Next, we discuss some of the special methodological concerns that make well-powered tests of billboards challenging. We then describe our experiment, which sought to overcome some of these challenges. The experimental design section describes the setting in which the study took place, the billboard message to be tested, the process of random assignment, and our method of assessing voters' expected exposure to the billboards on validated turnout. Across model specifications, turnout appears to respond modestly to signage, with somewhat larger effects among those who presumably were exposed repeatedly to the GOTV message. We conclude by discussing the implications of these results against the backdrop of other turnout experiments in high-salience elections and 2020 in particular.

Competing Theoretical Perspectives

The empirical literature on billboard advertising originated in the 1960s and has largely relied on lab-based tests of advertising content or surveys assessing the link between consumer product attitudes and exposure to outdoor advertisements (Danaher, 2017; Donthu *et al.*, 1993; Zekiri, 2019). This literature seems to support two broad empirical propositions. The first is that billboards attract the attention of a large proportion of passers-by, a feature that at times sparks controversy insofar as billboards draw motorists' attention away from the road (Belyusar *et al.*, 2016). The second is that billboards seem to impart basic information, such as awareness of brands or upcoming events (Fortenberry *et al.*, 2010).

Prevailing theories about why billboards may change behaviors — especially consumer purchases — stress their ability to attract viewers' attention, pique their interest, and create a desire for the advertised product or service (Fortenberry and McGoldrick, 2020). Lab- and survey-based experiments seem to show support for these effects (Kronrod and Huber, 2019), and interviews with advertising firms and organizations confirm that billboards are purchased with these causal effects in mind (Taylor *et al.*, 2006). The presumption seems to be that billboard advertising is both effective and increasingly attractive as a medium that can reach audiences that would otherwise be missed by online ads. Even critics of billboards, who point out that they often encourage low-income communities to consume alcohol or junk food (Yancey *et al.*, 2009), implicitly agree about the effects of billboard advertising on commercial sales.

Nevertheless, convincing demonstrations of outdoor advertising's behavioral effects are few in number, despite decades of pleas for field experimentation (Woodside, 1990). As summarized in Table 1, a small number of geographically focused studies have quasi-randomly manipulated the timing and placement of billboards to encourage behaviors such as seeking health care (Fortenberry *et al.*, 2010), utilizing public services (Calderwood and Wellington, 2015), or

Table 1: Prior field research on the effectiveness of billboard advertising.

	Experimental	Outcome	Finding
Panagopoulos and Ha (2015)	✓	Voting	0
Minkoff and Mann (2020)	✓	Voting	+
Bhargava and Donthu (1999)	✓	Recall	+
Fortenberry *et al.* (2010)		Recall	+
Papas *et al.* (2004)		Recall	+
Nebenzahl and Hornik (1985)	✓	Recall	0
Chaney *et al.* (2004)		Recall	Mixed
Frison *et al.* (2014)		Sales	Mixed
Berkowitz *et al.* (2001)		Sales	+

visiting local tourist sites (Bhargava and Donthu, 1999), producing mixed results. Another strand of research uses time-series analysis to assess the elasticity of sales in the wake of fluctuations in outdoor advertising. This line of research, too, has generated ambiguous results that seem sensitive to modeling assumptions (Berkowitz *et al.*, 2001; Bhargava *et al.*, 1994; Frison *et al.*, 2014). To our knowledge, the most recent review of this literature is Danaher (2017), which mentions only three field-based studies of billboards, none of which uses random assignment.

Largely missing from the consumer behavior literature are theories about why billboard advertising might fail to produce meaningful behavioral effects. The literature on voter mobilization, which evaluates a wide range of tactics in addition to signage, is instructive insofar as it relies on much stronger field experimental designs that often suggest weak effects. First, the mobilization literature casts doubt on whether even get-out-the-vote messages that command voters' attention are sufficient to raise turnout. Live phone calls from telemarketing firms that remind people to vote in an upcoming election typically produce modest effects, even when researchers focus solely on "compliers" who actually answer the phone (Arceneaux *et al.*, 2006). Second, few messages pique voters' interest in an upcoming election or are sufficiently memorable to influence voting behavior days or weeks after they are received (Green and Gerber, 2019, chap. 11). Important exceptions include personal encouragements from friends (Teresi and Michelson, 2015) or forceful expressions of social norms (Gerber *et al.*, 2008), which are more likely to inspire an intention to vote and to be remembered long enough to influence behavior. Although, as mentioned above, some previous studies of signage and turnout have generated positive estimates, no study has isolated positive effects with precision.

Putting these theoretical perspectives together leads to conflicting predictions. Militating in favor of positive effects is the notion that exposure to signage in the midst of an election campaign, especially during a period when many states allowed some form of early voting, encourages viewers to promptly express their boosted motivation to vote through actual behavior, just as billboards direct hungry travelers to nearby restaurants. On the other hand, it may be that in the context of a very high-salience election, signage fails to impart additional motivation to the relatively small segment of the electorate that would not otherwise vote. Given the dearth of previous studies assessing the effects of billboards in general, election-related billboards in particular, and signage in high-salience contexts, the present study fills an important gap in the research literature.

Experimental Design

Our study builds on the design developed by Minkoff and Mann (2020) insofar as we randomly assign potential billboard locations to GOTV signage. The primary difference between our design and theirs is that they randomly selected billboard locations within a small set of metro areas. Within each target area, Minkoff and Mann (2020) overlaid a grid of 3×3 mile cells. Cells were then randomly assigned to treatment and control groups, and only billboards located within treatment cells were purchased. Since some treated billboards were located near control billboards, spatial spillover effects are a concern, as some people who lived in control locations may nevertheless have seen the billboards when traveling. By contrast, our design randomly assigns different *metro areas* to treatment or control. Our design is, therefore, less susceptible to spillover effects. The next section describes the details of our experimental design.

Selecting Eligible Billboards

The experimental billboard campaign was carried out during the final four weeks before the 2020 US presidential elections. Three advertising companies furnished lists of available vinyl and digital billboards, and we selected a final list of billboards based on the following criteria. First, in order to reduce the chances that our billboards would have a partisan effect, only boards in zip codes in which Blacks and Hispanics comprised at least 25% of the population were retained. Second, billboards that cost more than $7,000 were dropped due to budget constraints. Third, in order to make it easier to assess effects on voter turnout, we focused solely on billboards close to residential or commercial areas, eliminating boards along freeways or in industrial areas.

Like Minkoff and Mann (2020), we selected billboards whose primary audience consists of individuals who live in close proximity to them. This targeting strategy is important when we model and estimate individuals' likely exposure to the billboard. Since we retained boards that are in or beside residential areas and excluded boards that are along highways, we expect the level of exposure to a billboard to decrease sharply as the distance between one's residence and the billboard increases. Below we describe our method for empirically assessing the functional relationship between distance and expected exposure.

Random Assignment of Treatment

We assign billboards to treatment or control, where a treated billboard contains a GOTV message. In order to ensure geographic buffers between treatment and control areas, we conducted a two-level random assignment, first selecting metro areas and next selecting specific billboards within each chosen metro area. We randomized in four tranches, reflecting the fact that the boards and the funding to purchase them became available at different points in time. In effect, each tranche represents a distinct block within which randomization occurred.

In the first tranche, 182 vinyl boards from two firms were blocked based on the number of neighboring boards within 10 miles of each other. The first block contained 78 boards that had no other board within a 10 mile radius. The second block contained 38 clusters of boards, where each cluster contained a pair of boards within 10 miles of each other, yielding a total of 76 boards. The third block contained 8 clusters of boards, where each cluster had three or more boards within 10 miles of each other, yielding a total of 28 boards. The boards were block randomized, yielding 91 in treatment and 91 in control.[2] All treatment boards in the first tranche were displayed between October 7 and November 8.[3]

In the second tranche, a third firm provided a list of vinyl boards that were more than 20 miles away from any board in the first tranche. This yielded two boards, one of which was randomly assigned to treatment through the flip of a coin. The treatment board was displayed between October 12 and November 8.

In the third tranche, we purchased ad time for 10 clusters of 34 digital boards located in cities not yet covered by our vinyl boards. We randomly assigned half the clusters to treatment and the other half to control, yielding 14 boards in treatment and 20 boards in control. These digital billboards ran our ads between October 12 and November 3. In the fourth tranche, we purchased ad time for 19 clusters of 80 digital boards in cities without existing treatment or control boards between October 16 and November 3.

[2]Two of the treatment boards became unavailable after randomization. Both of these boards were in geographic clusters with more than one billboard. Because only a small fraction of billboards in the treatment group went untreated, we focus exclusively on the intent-to-treat effect.

[3]Vinyl installment took place over approximately three days (October 7 to 10). Boards in New Orleans were delayed an additional 2 days due to a hurricane.

Specifically, we blocked on the number of clusters per state, and whether the state was a battleground state. These criteria resulted in four blocks. The first block consists of five clusters of digital billboards within the state of Georgia. The second block consists of two clusters within the state of South Carolina. The third block consists of states with one cluster and that were considered battleground states (neither deep blue nor deep red) by predictit.org on October 15, 2020. The fourth and last block consists of non-battleground states with one billboard. We randomly assigned half of the clusters in each block to treatment, yielding 10 clusters of 42 boards in treatment and 9 clusters of 38 boards in control, depicted in Figure 1. Table 2 provides a summary of how many billboard clusters were assigned to the treatment and control in each tranche and block, and Table 3 shows the covariate profiles of registered voters whose addresses were within miles of the nearest treatment or control billboards.

Billboard Design

When designing our billboard message, we sought to convey a non-partisan, GOTV message comparable to the Minkoff and Mann (2020) treatment and sensitive to the unique features of the 2020 election cycle. Minkoff and Mann (2020) deployed red billboards that included the word "vote" in bold, the date of the election, and reference to the sponsoring organization, Vote.org. Because early and mail-in voting were expected to be far more common in the 2020 election, we chose to emphasize that citizens vote *by* Election Day. We also

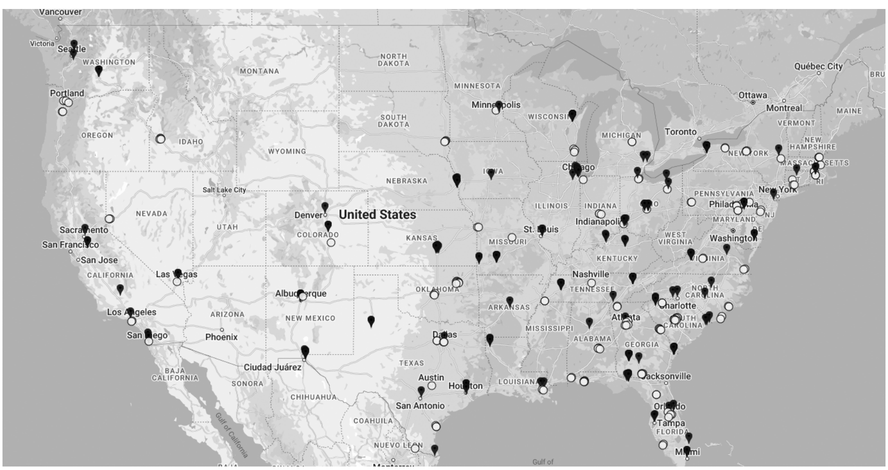

Figure 1: National map of billboard locations. (Treatment boards are black pointers, control are white circles.)

Table 2: Summary of billboard assignments.

Tranche	Dates available	Board type	Block	Treatment clusters (boards)	Control clusters (boards)
1	10/7/2020–11/8/2020	Vinyl	1	39 (39)	39 (39)
			2	19 (38)	19 (38)
			3	4 (14)	4 (14)
2	10/12/2020–11/8/2020	Vinyl	1	1 (1)	1 (1)
3	10/12/2020–11/3/2020	Digital	1	5 (20)	5 (14)
4	10/16/2020–11/3/2020	Digital	1	3 (10)	2 (9)
			2	1 (3)	1 (9)
			3	3 (12)	3 (9)
			4	3 (13)	3 (15)
Total				78 (150)	77 (148)

Table 3: Covariate profiles of registered voters whose address is closer to either a treatment or a control billboard.

	Control			Treatment		
	N	Mean	SD	*N*	Mean	SD
Voted in 2016	8,630,029	0.61	0.49	9,824,414	0.59	0.49
Voted in 2012	8,630,029	0.53	0.50	9,824,414	0.52	0.50
Probability white*	8,622,729	0.63	0.36	9,817,168	0.57	0.38
Age	8,581,263	46.84	18.00	9,813,636	47.11	18.18

A 5-mile radius was selected to determine how many voters were included.
*TargetSmart provided modeled estimates of race that come in the form of a posterior probabilities. This table reports the average probability that a subject is non-Hispanic White, among those in the treatment or control groups.

added a simple phrase, "make your voice heard," in an attempt to strengthen the advertisement's impact on behavior. Finally, we chose a white-on-black color scheme to eliminate any partisan connotations of colors such as red or blue. The billboard design is shown in the Appendix.

Confirming Compliance with Assigned Treatment

The firms that posted the treatment billboards provided digital images confirming their deployment.[4] Only one of the vinyl control billboards displayed a GOTV message (by Vote.org). None of the other control billboards featured an election-related message. Given the small extent of non-compliance with assigned treatment, we make no attempt to correct for it statistically.

Modeling Exposure to Billboards

Define the *true exposure* $E_{ij}^{(\star)} \in \{0, 1, 2, \dots\}$ to be the number of times subject i was within reading distance[5] of billboard j during the study period. Define also the *treatment status* Z_j of board j to denote whether it displayed a GOTV message ($Z_j = 1$) or was part of the control group ($Z_j = 0$). The *true treatment dosage* $D_i^{(\star)}$ of subject i is then the total number of times they were exposed to a treated board,

$$D_i^{(\star)} = \sum_j Z_j E_{ij}^{(\star)}. \tag{1}$$

In our setting, we do not directly observe subjects' true exposures to billboards and thus do not observe their true treatment dosages. We instead rely on two different modeling approaches to assess the causal effects of GOTV billboards on voting behavior: (1) a *coarse-grained approach* that compares aggregate voting rates between geographic regions which contained different numbers of treated billboards (see the section "Coarse-Grained Model of Voting Rates"), and (2) an *individual-level approach* that relies on an imputed measure of subjects' treatment dosage (see the section "Individual-Level Model of Voting Outcomes with Imputed Exposures").

Coarse-Grained Model of Voting Rates

Since we do not observe a subject's true exposure, we first consider instead their *coarse exposure* at a given radius r, $\tilde{E}_{ij}^{(r)} \in \{0, 1\}$, which denotes whether

[4]In the case of digital boards, we were unable to document the types of advertisements that were running on the 62 control boards leading up to the election. Several advertisements run on the same digital board during any given period, and the flight length for advertisements is much more variable and often shorter than traditional vinyl boards.

[5]The industry standard viewing distance ranges between 500 and 1000 ft. In our analysis, we consider "within reading distance" to mean within 200 m (656 ft.) of a billboard.

billboard j is within r miles of subject i's registered address:

$$\tilde{E}_{ij}^{(r)} = \mathbf{1}(\text{subject } i \text{ lives within } r \text{ miles of billboard } j). \tag{2}$$

The database we use to observe subjects' voting outcomes also lists their registered address with a corresponding latitude and longitude — so, along with the latitude and longitude of each billboard, we can compute the Euclidean distances between all (subject i, billboard j) pairs, which allows us to compute Eq. (2) for a given radius r. As radii diminish (e.g., $r = \frac{1}{2}$ mile), subjects will be coded as exposed to fewer billboards.

Define $\tilde{\boldsymbol{E}}_i^{(r)} = (\tilde{E}_{i1}^{(r)}, \ldots, \tilde{E}_{im}^{(r)})$ to be the binary vector of coarse exposures for subject i across all m billboards. At a given radius r, there is a fixed subset $\mathcal{S}^{(r)} \subseteq \{0,1\}^m$ of values that $\tilde{\boldsymbol{E}}_i^{(r)}$ can feasibly take, since there are only certain combinations of billboards near enough to each other to all be within an r-radius of subject i's registered address. As r increases, the size of the set $\mathcal{S}^{(r)}$ increases.

In this section, the units of analysis correspond to the elements of $\mathcal{S}^{(r)}$, each of which define a unique geographic region within which residents are exposed to a unique combination of billboards. We will use $s \in \{1, \ldots, |\mathcal{S}^{(r)}|\}$ to index the set of geographic units for a given radius r, and $\boldsymbol{e}_s = (e_{s1}, \ldots, e_{sm}) \in \mathcal{S}^{(r)}$ to denote the unique combination of billboards corresponding to geographic unit s. See Figure 2 for a simple example of how geographic regions might overlap. Note that two regions s and s' never overlap geographically even though the unique combinations of billboards they correspond to may overlap — i.e., the indices where $\boldsymbol{e}_s$ and $\boldsymbol{e}_{s'}$ are 1 may overlap.

Of all n subjects, the number residing within geographic unit s is n_s, which can be defined simply in terms of the subject-specific exposures:

$$n_s = \sum_{i=1}^{n} \mathbf{1}(\tilde{\boldsymbol{E}}_i^{(r)} = \boldsymbol{e}_s). \tag{3}$$

All n_s subjects within region s receive the same *treatment dosage* D_s, defined as the number of treated billboards among those in the unique combination corresponding to region s:

$$D_s = \sum_{j=1}^{m} e_{sj}\, Z_j. \tag{4}$$

Finally, subjects residing in unit s all have the same *treatment dosage propensity*, $\rho_s > 0$,

$$\rho_s = \mathbb{E}_{\boldsymbol{Z}}\left[D_s\right] = \tfrac{1}{2} \sum_{j=1}^{m} e_{sj}, \tag{5}$$

Figure 2: Example of 1-mile radii around treatment and control boards (treatment is striped).

where the expectation is with respect to the distribution implied by all possible randomizations of the treatment status vector $\boldsymbol{Z}$. This propensity may be inferred from the experimental design since billboards are assigned to treatment with known probability. The identification strategy hinges on comparing turnout in geographic units that have the same propensity for treatment (i.e., have the same number of billboards that are eligible for treatment assignment with a given probability) but, due to random assignment, are exposed either to treated billboards or untreated billboards. Thus, conditioning on propensity is crucial for causal identification.

Note that the number of subjects n_s, treatment dosage D_s, and ρ_s all implicitly depend on the radius r, since s indexes into the set of regions $\mathcal{S}^{(r)}$, which varies by r. To be explicit about this, we will write them as $n_s^{(r)}$, $D_s^{(r)}$, and $\rho_s^{(r)}$.

We express the average marginal effect of a one-unit increase in dosage $\alpha_1^{(\text{coarse})}$ using the following structural equation:

$$\mathbb{E}[\bar{Y}_s^{(r)}] = \alpha_0 + \alpha_1^{(\text{coarse})}\, D_s^{(r)} + \alpha_2\, \rho_s^{(r)} + \boldsymbol{\alpha}_3^\top \boldsymbol{X}_s \tag{6}$$

where $\bar{Y}_s^{(r)}$ is the voting rate in 2020 of region s, and $\boldsymbol{X}_s$ are covariates about s, such as voting rates in prior elections. We estimate the parameters of this model using weighted OLS, setting the weights to n_s so that more populous clusters receive more weight, as would be the case if we were estimating an individual-level average treatment effect.[6] Results are reported in Table 4, where randomization inference is used to obtain a p-value associated with $\hat{\alpha}_1^{(\text{coarse})}$.

Individual-Level Model of Voting Outcomes with Imputed Exposures

As an alternative to the coarse-grained analysis, we also consider an individual-level analysis. We obtained data from The Center for New Data, a non-partisan non-profit that was founded in 2020. Using their national cell phone location data, we were able to observe whether and how often individuals were in the vicinity of our treatment and control billboards.

While the ideal approach would be directly tracking each subject's actual exposure to billboards, our data usage agreement and IRB protocol prevent us from directly linking voter turnout data to individual cell phones. Thus, we develop an individual-level analysis that relies on an imputed measure of subjects' exposures. A subject's *imputed exposure* $\hat{E}_{ij} > 0$ is the expectation of their true exposure $E_{ij}^{(\star)}$ under a parametric model $P_\theta(E^{(\star)} \mid \boldsymbol{X})$ that we train to predict true exposures from demographic covariates $\boldsymbol{X}_i$:

$$\hat{E}_{ij} = \mathbb{E}_\theta[E_{ij}^{(\star)} \mid \boldsymbol{X}_i]. \tag{7}$$

We train this model using an ancillary dataset of cell phone GPS pings that allows us to observe the ground truth exposures for a subset of individuals; more details on this model are given below. We use these imputed exposures to define subjects' *imputed treatment dosage* $\hat{D}_i \geq 0$, defined as

$$\hat{D}_i = \sum_j \hat{E}_{ij}\, Z_j. \tag{8}$$

[6]Although the design involves clusters of unequal size, which potentially jeopardizes the unbiasedness of the regression estimator, the large number of clusters means this bias is unlikely to materially affect the results.

We further define a subject's *imputed treatment dosage propensity* $\hat{\rho}_i > 0$ as the expectation of their treatment dosage under the randomization distribution

$$\hat{\rho}_i = \mathbb{E}_{\boldsymbol{Z}}[\hat{D}_i] = \tfrac{1}{2} \sum_{j=1}^{m} \hat{E}_{ij}. \tag{9}$$

When checking the robustness of our results, we verify that our substantive conclusions do not change depending on the minimum value of $\hat{\rho}_i$.

We then express the average marginal effect of a one-unit increase in individual-level dosage $\alpha_1^{(\text{indiv})}$ using the following equation:

$$\mathbb{E}[Y_i] = \alpha_0 + \alpha_1^{(\text{indiv})}\, \hat{D}_i + \alpha_2\, \hat{\rho}_i + \boldsymbol{\alpha}_3^\top \boldsymbol{X}_i \tag{10}$$

where $Y_i \in \{0, 1\}$ is subject i's 2020 voting outcome and $\boldsymbol{X}_i$ are demographic covariates including subject i's voting outcomes in prior elections. We fit this model using OLS; the results are given in Table 5. In order to make the results more readily interpretable, we estimate a similar regression in which dosage is an indicator variable ($1 = \text{treatment}$, $0 = \text{control}$), stratifying by quartiles of expected exposure. These regressions (see Appendix Tables 14 and 15) provide a sense of whether the average effect of the treatment rises as average dosage rises.

Details on imputed exposures

The imputed exposure ($\widehat{E}_i$) is coded as the expected number of days during which an individual i was near a treated billboard.[7] We obtain this imputation by modeling how likely an individual is to have been in proximity to one of our billboards during the period of treatment. We obtained location data from the Center for New Data on both individuals known to live within a five-mile radius of our billboards and individuals whose phones "pinged," or provided a location stamp within 200 meters of a board at some point during our experiment. We then matched pings to our catchment areas, tracking who among the cell phone sample pinged near a board. Exposure to most boards dropped exponentially with distance. We, therefore, use the log of distance as an independent variable and fit a model for each board with exposure as the dependent variable. Thus, for each voter we impute an exposure based on the

[7] We chose to model expected number of days for two reasons. First, it makes sense to differentiate individuals who passed by a board once from individuals who passed almost every day; seeing a board requires both proximity and attention, and the latter set of individuals are more likely to have noted the board. Second, considering number of days improves model precision; the alternative of using a binary indicator underestimates the effect of distance on the probability of exposure.

model's fitted exposure to each billboard.[8] For almost all boards, predicted (imputed) exposure is negatively correlated with distance.[9] Our results remain unchanged substantively when we impose a single exposure model pooling over all of the locations, which shows the anticipated negative correlation between exposure and distance (Appendix Table 10).

Comparing the Coarse-Grained and Individual-Level Estimators

Both estimation approaches attempt to gauge the causal effects of billboards on voting and can be viewed as complementary robustness checks. The two estimators can also be viewed as targeting slightly different causal effects. To take the simplest case in which voters reside within radius r of just one billboard, the coarse-grained approach estimates the average effect of being within the specified catchment area of a treatment billboard, whereas the individual approach assesses the average effects of voters' proximity to the treated billboard (since proximity is what drives the expected exposure metric that is used to scale the effect). Put somewhat differently, the coarse-grained approach compares geographic units within the same radius of local billboards; the individual-level analysis compares voters with the same expected exposure based on travel patterns. A series of additional robustness checks assess whether the results change when individual-level data are modeled using geographic proximity to billboards instead of cell phone data.

Results

Our analysis begins by considering billboards' effects on voter turnout rates at the geographic-region level. For a given radius r (see the section "Coarse-Grained Model of Voting Rates"), the set of regions is defined according to the locations of the billboards $\mathcal{S}^{(r)}$. Each region encompasses registered voters' addresses, and the outcome variable is the voting rate for each cluster of voters. We consider different radii, starting with a radius of 0.5 miles and working up to 5 miles. The Appendix repeats this analysis, with similar results, using the

[8] In our pre-analysis plan, we initially envisioned a more complex machine learning model to predict exposure, as well as a larger radius. After viewing cell phone movement patterns, it became clear that the probability of exposure drops rapidly and is expected to be close to zero after about 5 miles, so we restricted our focus to that distance. We also explored the predictive contributions of other available covariates, but none significantly improved predictive capacity in a consistent and theoretically reasonable way across boards. Thus, we opted for a simple model based on distance.

[9] In a few exceptional cases, exposure is weakly positively correlated with distance. We manually inspected each of these cases and found that features of geography explain the exceptions. For example, one board with this pattern was placed by a central bridge within a city; as a result, commuters were more likely to see the board than were locals.

raw number of votes cast in each region as the outcome variable (as opposed to the voter turnout rate).

In this analysis, all subjects within region s receive the same *treatment dosage* $\tilde{D}_s^{(r)}$ (see Eq. (4)). At a radius of $r = 0.5$ miles, this ranges from zero, if a region, or *catchment area*, includes no treated billboards, to three, if it included three treated billboards.[10] Of the 305 catchment areas, 49.5% were assigned zero treated billboards, 44.9% were assigned one treated billboard, 4.6% two, and the remainder three. The first column of Table 4 reports the estimated average marginal effects of treatment dosage on each area's turnout rate. The point estimate is found to be weakly positive (0.004, or four tenths of a percentage point), and randomization inference renders a one-tailed p-value of 0.28.

In order to get an intuitive feel for the results, it is helpful to look at Figure 3, which plots the residuals of the dependent variable (when regressed on all covariates) against the residuals of dosage (when regressed on all covariates). For each radius, the slope of the regression line that passes through these (weighted) points exactly reproduces the estimated marginal dosage effect reported in Table 4. These regression lines suggest that dosage has a modest effect on turnout rates. The standard error of the estimated marginal effect is small as well, due to the fact that past voting rates are strong predictors or voter turnout in 2020, with R^2 values of 0.77 and up.

At a radius of one mile, the sign of the estimated effect remains weakly positive and statistically indistinguishable from zero (randomization inference $p = 0.23$). This pattern of estimated average marginal effects hovering just above zero persists as we expand the radius, which increases the effective number of catchment areas. When the radius is 5 miles, the point estimate is 0.003. Taken at face value, this estimate implies that a maximal treatment dosage (i.e., four nearby billboards) raised turnout in a catchment area by 1.2 percentage points. However, Figure 3 offers a visual reminder that the relationship between dosage and voting rates is subtle, regardless of radius size. This impression is confirmed by the randomization inference p-values in Table 4, which never approach conventional levels of significance, regardless of radius.

Similar findings emerge when we change the modeling strategy so that the analysis focuses on individual voters rather than geographic clusters (see see the section "Individual-Level Model of Voting Outcomes with Imputed Exposures"). Table 5 reports the results from three regressions. The first analyzes all registered voters who live within 5 miles of a billboard that was eligible for treatment. In addition to controlling for voter turnout in the preceding four elections, the regression controls for *dosage propensity* (see Eq. (5)) across all possible random assignments. Adjusting for this covariate is crucial because we want to know the effect of treatment dosage for people who

[10] At $r = 0.5$ miles, no clusters overlap.

Table 4: Marginal effect of treatment dosage on a geographic region's voter turnout rate, by radius.

	Dependent variable:				
	2020 Turnout rate				
	0.5 Mile	1 Mile	2 Miles	3 Miles	5 Miles
	(1)	(2)	(3)	(4)	(5)
Treatment (Dosage)	0.004	0.004	0.003	0.003	0.003
	(0.006)	(0.004)	(0.003)	(0.003)	(0.002)
2018 Turnout rate	0.750	0.638	0.604	0.585	0.586
	(0.062)	(0.055)	(0.051)	(0.049)	(0.045)
2016 Turnout rate	0.583	0.757	0.821	0.806	0.853
	(0.081)	(0.073)	(0.067)	(0.065)	(0.063)
2014 Turnout rate	0.072	0.073	0.076	0.080	0.048
	(0.065)	(0.055)	(0.047)	(0.043)	(0.040)
2012 Turnout rate	−0.412	−0.487	−0.536	−0.524	−0.555
	(0.066)	(0.059)	(0.053)	(0.050)	(0.048)
Treatment propensity = 0.5	0.004	0.020	0.017	0.013	0.066
	(0.060)	(0.271)	(0.139)	(0.027)	(0.251)
Treatment propensity = 1	−0.018	0.004	0.004	0.006	0.057
	(0.062)	(0.271)	(0.139)	(0.027)	(0.251)
Treatment propensity = 1.5		0.004	0.003	0.002	0.049
		(0.272)	(0.139)	(0.028)	(0.251)
Treatment propensity = 2			0.001	−0.00001	0.046
			(0.141)	(0.030)	(0.251)
Constant	0.150	0.128	0.137	0.152	0.104
	(0.064)	(0.272)	(0.140)	(0.031)	(0.252)
RI p-value	0.2751	0.2258	0.1842	0.1877	0.1984
Pct geoclusters exposure >1	0.05	0.121	0.203	0.227	0.244
Observations	301	354	448	528	624
R^2	0.791	0.818	0.818	0.792	0.772
Adjusted R^2	0.786	0.814	0.814	0.789	0.769

Treatment dosage: clusters may be exposed from 0 to 3 treated boards in the 0.5-mile radius; 0 to 3 treated boards in the 1-mile radius; 0 to 4 treated boards in the 2-mile radius; 0 to 5 treated boards in the 3-mile radius; and 0 to 4 treated boards in the 5-mile radius.
RI p-values: the p-values derived by randomization inference for each regression model are calculated based on a large number of possible block and cluster assignments of billboards to treatment.
Pct geoclusters exposure >1: percentage of geographic clusters exposed to more than one billboard for a given radius.
Observations in each regression are restricted to geographic clusters with more than 100 registered voters. All models weight by total number of registered voters in each geographic cluster.

otherwise have the same propensity to encounter the experimental billboards. The distribution of imputed dosage ranges from just above 0 to 17.55, with a mean of 0.18 and standard deviation of 0.5. The first column of Table 5 reports a weakly negative relationship between treatment dosage and turnout,

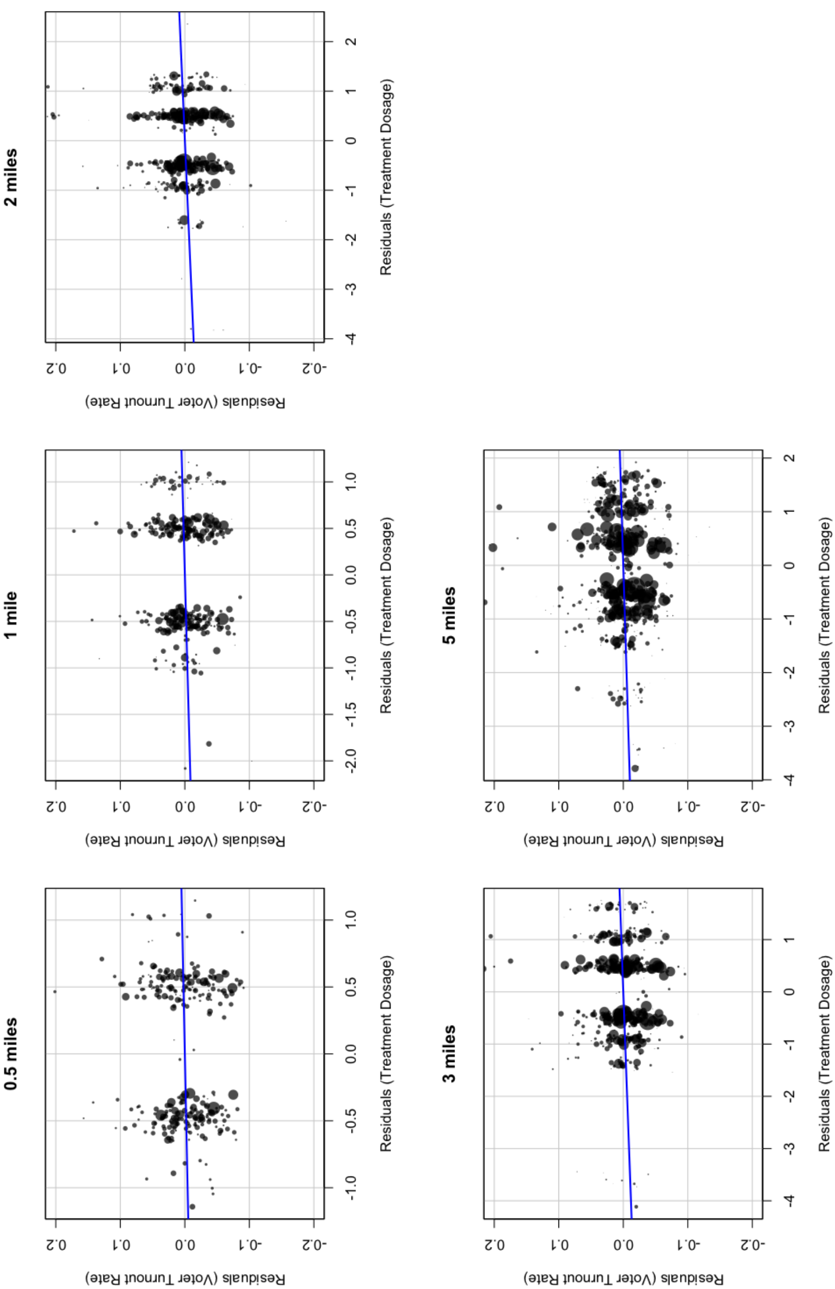

Figure 3: Visualization of marginal effects of dosage on turnout rates, by radius.

Table 5: Marginal effect of imputed treatment dosage on individual-level voter turnout, overall and by whether subjects are registered in battleground states.

	Dependent variable:		
		Voted in 2020	
	Overall	Battleground	Non-battleground
	(1)	(2)	(3)
Treatment (Dosage)	−0.002	−0.007	0.003
	(0.004)	(0.005)	(0.005)
Exposure to all boards	−0.005	−0.006	−0.006
	(0.004)	(0.005)	(0.005)
Voted 2018	0.305	0.320	0.293
	(0.004)	(0.005)	(0.005)
Voted 2016	0.202	0.203	0.199
	(0.003)	(0.004)	(0.004)
Voted 2014	0.029	0.030	0.026
	(0.002)	(0.002)	(0.002)
Voted 2012	0.054	0.058	0.051
	(0.002)	(0.003)	(0.002)
Constant	0.386	0.360	0.409
	(0.008)	(0.010)	(0.012)
RI p-value	0.891	0.822	0.771
Observations	18,454,443	8,047,875	10,406,568
Clustered SEs	✓	✓	✓
R^2	0.291	0.298	0.284
Adjusted R^2	0.291	0.298	0.284

The variable *Exposure to all boards* is the predicted number of days that the individual was exposed to at least one board, either treated or control. It has a mean of 0.397 with a standard deviation of 0.717.
The variable *Treatment (Dosage)* is the predicted number of days that the individual was exposed to at least one treated board. It has a mean of 0.199 with a standard deviation of 0.543.
These regressions include only individuals living within a maximum radius of five miles from any billboard. Standard errors were clustered by billboard cluster. No weights were included for the individual-level regressions.

implying that turnout declines by 0.2 percentage points per additional imputed exposure. The estimated clustered standard error is 0.4 percentage points, so the apparent effect is not statistically distinguishable from zero.

The basic pattern of results remains largely unchanged when, per our pre-analysis plan, we partition the data according to whether voters reside in presidential battleground states. The estimated dosage effect among those living in battleground states is negative but statistically indistinguishable from zero (−0.007, SE=0.005). The estimated effect in non-battleground states is weakly positive (0.003, SE=0.005). Although the apparent effect in non-battleground states is, as expected, larger than in battleground states, the two estimated effects are not significantly different from each other. One post hoc comparison (suggested to us by a reader of an earlier draft of this essay) considers whether digital boards are more influential than vinyl boards, on the grounds that digital boards are far more noticeable, particularly at night. As shown in the Appendix (Section 5.12), digital ads do not appear to be especially effective.

What accounts for the discrepancy between the positive estimates retrieved by the coarse-grained analysis and the negative estimate rendered by the pooled individual-level analysis? As shown in the Appendix, the coarse-grained results obtain under a variety of modeling approaches. For example, a difference-in-difference analysis within randomization tranches produces a similar pattern, with larger estimates among those living within a half-mile of a treated billboard (Appendix Table 7). The same goes for other models that ignore dosage and simply use a binary indicator for exposure to treatment (Appendix Section 5.5), and when controlling for a polynomial in expected exposure (Appendix Table 6); again, the only substantial estimates are among those residing within a half-mile of a treated billboard. Interestingly, when we revisit the individual-level analysis in the first column of Table 5 but vary the radius (Appendix Table 12), we again find the largest point estimates for the half-mile radius. In short, a variety of robustness checks suggest that results are sensitive to some modeling nuances but not others. It seems not to matter whether the data are aggregate or individual, or whether the regression model scales dosage using cell phone data, using a quadratic function of distance to the nearest billboard, or using a simple binary indicator. To the extent that the results vary, they change according to the size of the radius around each billboard. The smaller the radius, the larger the estimated effect. But even when the radius is one-half mile, the estimates fall short of conventional 0.05 levels of statistical significance using a one-tailed test.

The remaining question is whether the experimental billboards increased voter registration. Using the number of new registrants per geographic cluster as the outcome, we estimate this marginal effect for a range of cluster sizes in Appendix Table 19. Although the experiment is well-powered to detect effects on the order of a few dozen new registrants per cluster, the estimated effects are all weakly negative and statistically indistinguishable from zero.

Conclusion

When launching this experiment, we had reason to suspect that billboard advertising might work. The largest prior experimental evaluation, conducted in low salience elections, found that billboard signage raised turnout among voters residing nearby (Minkoff and Mann, 2020). The broader literature on billboard advertising, although largely reliant on evaluation methods other than field experiments, tends to be optimistic about the effects of signage on consumer behavior. One recurrent theme in that literature is that billboard advertising has the advantage of conveying a widely viewed message outside the digital and television context, which are cluttered with ads.

Our experiment offers an unusually precise reading of the average treatment effect of billboard advertising. The number of metro areas in our study is an order of magnitude larger than in any previous study, and the accuracy with which behavioral outcomes — voter registration and turnout — are measured set the study apart from prior research that has often relied on survey responses. Although modeling the statistical effects of billboard advertising on outcomes is not straightforward, given the inherent ambiguity in how geographically diffuse the effects of advertising may be, all of the pre-specified models we applied point to the same conclusion. Most point estimates are weakly positive; some are weakly negative; and all are statistically indistinguishable from zero.[11]

This finding is subject to three competing interpretations. The most cautious interpretation, which takes note of the fact that the weakest estimates we obtain are very close to zero, calls into question the efficacy of billboard advertising on the grounds that the current experiment is much larger than previous GOTV billboard studies and both larger and more rigorously designed than the extant billboard literature on topics other than political participation. The average treatment effects of billboards may be so small as to elude detection by even a well-powered study.

A less skeptical interpretation takes note of the fact that some of our point estimates are large enough to be consequential. Although statistically insignificant at conventional levels, high levels of dosage may produce meaningful

[11]In the spirit of leaving no stone unturned, we explored other models that attempt to zero in on effects among those whose expected exposure was highest. We divided voters into quartiles based on their expected exposure and estimated the effect of treatment assignment (a binary variable), controlling for the usual covariates, such as expected exposure). The results presented in Appendix Tables 14 and 15, using 5-mile and 1-mile radii, respectively, are equivocal. We find weak estimates using the larger radius; the smaller radius renders a 2 percentage point effect in the third quartile but only a 0.3 percentage point effect at the highest dosage. Appendix Table 13 hints at positive effects when the analysis focuses solely on voters living within one-half mile of the billboards; expanding the radius makes these estimates weakly negative. The largest estimate is a 1.7 percentage point increase per exposure in non-battleground states within a one-half mile radius (Appendix Table 21). The exploratory analysis that turned up the largest effect is shown in Appendix Table 18, which restricts attention to states that discouraged early and by-mail voting. Here we find weak effects when including voters living more than one mile from the closest billboard but implausibly large effects if we restrict attention to voters living within one-half mile. Future research should investigate whether billboards are indeed especially effective in states with traditional voting systems.

effects. The coarse-grained analysis indicates that turnout increased by 1.2 percentage points among those living within one-half mile of three billboards.[12] Similarly, the individual-level analysis in column 1 of Appendix Table 13 indicates that those registered voters living within one-half mile of at least one eligible billboard experienced a 1.3 percentage point higher rate of turnout where GOTV messages were deployed. If these estimates are to be trusted, they would imply that billboard messages have meaningful behavioral effects among those exposed to them repeatedly, in which case the cost effectiveness of this tactic hinges crucially on voter population density and advertising costs.

An intermediate interpretation holds that billboards, though often effective, are relatively ineffective in the context of a high-salience election, such as a presidential election. This particular election was especially salient, attracting a higher percentage of eligible voters than any American election in more than a century. Even rates of turnout in non-battleground states were higher in 2020 than they had been in prior elections. This interpretation is supported by recent meta-analyses of 250 experiments showing that high-salience elections diminish the effectiveness of GOTV tactics such as direct mail, canvassing, and SMS messages by 30%–50% (Mann and Haenschen, 2022). Another meta-analysis of direct mail that included high- and low-salience elections from outside the United States (Fortier-Chouinard *et al.*, 2022) came to the same conclusion. This conclusion is further corroborated by other studies that track the effects of specific interventions deployed prior to and during the 2020 general election. Daniels *et al.* (2021), for example, find that COVID-era party materials designed to encourage home voting celebrations around mail-in balloting had substantial effects during 2020 primaries but no apparent effect in the general election. Similarly, Schein *et al.* (2021) find that friend-to-friend texting, which produced large turnout effects during the 2018 midterms, had much weaker effects in the 2020 general election. It may be that, at the margin, the non-voters who remained to be mobilized by billboards in the final weeks of the 2020 election were difficult to move by any method. It remains to be seen whether future tests of billboard messaging find its mobilizing effects to be substantially larger in low salience contests or on forms of political participation other than voting.

References

Arceneaux, K., A. S. Gerber, and D. P. Green. 2006. "Comparing Experimental and Matching Methods Using a Large-scale Voter Mobilization Experiment". *Political Analysis*. 14(1): 37–62.

[12]See the Appendix for other analyses that come to the same conclusion — larger effects among those living within a half-mile of a treated billboard — without attempting to measure the number of proximal billboards or to use cell phone data to predict exposure.

Belyusar, D., B. Reimer, B. Mehler, and J. F. Coughlin. 2016. “A Field Study on the Effects of Digital Billboards on Glance Behavior During Highway Driving”. *Accident Analysis & Prevention*. 88: 88–96.

Berkowitz, D., A. Allaway, and G. D’souza. 2001. “The Impact of Differential Lag Effects on the Allocation of Advertising Budgets Across Media”. *Journal of Advertising Research*. 41(2): 27–7.

Bhargava, M. and N. Donthu. 1999. “Sales Response to Outdoor Advertising”. *Journal of Advertising Research*. 39(4): 7–18.

Bhargava, M., N. Donthu, and R. Caron. 1994. “Improving the Effectiveness of Outdoor Advertising: Lessons from a Study of 282 Campaigns”. *Journal of Advertising Research*. 34(2): 46–56.

Blodgett, J. and B. Lofy. 2008. *Winning Your Election the Wellstone Way: A Comprehensive Guide for Candidates and Campaign Workers*. University of Minnesota Press.

Calderwood, K. A. and W. J. Wellington. 2015. “Using Roadside Billboard Posters to Increase Admission Rates to Problem Gambling Services: Reflections on Failure”. *Health Promotion Practice*. 16(4): 510–22.

Chaney, I. M., K.-H. Lin, and J. Chaney. 2004. “The Effect of Billboards within the Gaming Environment”. *Journal of Interactive Advertising*. 5(1): 37–45.

Dale, A. and A. Strauss. 2009. “Don’t Forget to Vote: Text Message Reminders as a Mobilization Tool”. *American Journal of Political Science*. 53(4): 787–804.

Danaher, P. J. 2017. “Advertising Effectiveness and Media Exposure”. In: *Handbook of Marketing Decision Models*. Springer, 463–81.

Daniels, A., S. DeMora, S. Hayes, and M. Michelson. 2021. *Party at the Mailbox: Mobilizing Black Voters with Celebrations of Community*. Paper prepared for the Annual Meeting of the American Political Science Association.

Donthu, N., J. Cherian, and M. Bhargava. 1993. “Factors Influencing Recall of Outdoor Advertising”. *Journal of Advertising Research*. 33(3): 64–73.

Esteban-Casanelles, T. 2020. “The Effects of Exposure to Electoral Advertising: Evidence from Spain”. *Unpublished manuscript, Department of Economics, Columbia University*.

Fortenberry, J. L., J. K. Elrod, and P. J. McGoldrick. 2010. “Is Billboard Advertising Beneficial for Healthcare Organizations? An Investigation of Efficacy and Acceptability to Patients”. *Journal of Healthcare Management*. 55(2).

Fortenberry, J. L. and P. J. McGoldrick. 2020. “Do Billboard Advertisements Drive Customer Retention?: Expanding the “AIDA” Model to “AIDAR””. *Journal of Advertising Research*. 60(2): 135–47.

Fortier-Chouinard, A., M. A. Bodet, F. Gélineau, J. Savoie, and M. Ouimet. 2022. “Getting the Message Out: Why Mail-Delivered GOTV Interventions Succeed or Fail”. *PS: Political Science & Politics*: 1–8.

Frison, S., M. G. Dekimpe, C. Croux, and P. De Maeyer. 2014. "Billboard and Cinema Advertising: Missed Opportunity or Spoiled Arms?". *International Journal of Research in Marketing*. 31(4): 425–33.

Gerber, A. S., D. P. Green, and C. W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment". *American Political Science Review*. 102(1): 33–48.

Green, D. P. and A. S. Gerber. 2019. *Get Out the Vote: How to Increase Voter Turnout*. Brookings Institution Press.

Green, D. P., J. S. Krasno, A. Coppock, B. D. Farrer, B. Lenoir, and J. N. Zingher. 2016. "The Effects of Lawn Signs on Vote Outcomes: Results from Four Randomized Field Experiments". *Electoral Studies*. 41: 143–50.

Guzzetta, S. J. 2006. *The Campaign Manual: A Definitive Study of the Modern Political Campaign Process*. Political Publications.

Kronrod, A. and J. Huber. 2019. "Ad Wearout Wearout: How Time can Reverse the Negative Effect of Frequent Advertising Repetition on Brand Preference". *International Journal of Research in Marketing*. 36(2): 306–24.

Mann, C. and K. Haenschen. 2022. "I Never Meta-Analysis Like You Before: Voter Mobilization by Electoral Salience". *Paper presented at the Annual Meeting of the American Political Science Association.*

Minkoff, S. and C. Mann. 2020. "The Effect of Billboards on Turnout in 4 States: A Randomized Experiment". *Paper presented at the Annual Meeting of the American Political Science Association.*

Nebenzahl, I. and J. Hornik. 1985. "An Experimental Study of the Effectiveness of Commercial Billboards in Televised Sports Arenas". *International Journal of Advertising*. 4(1): 27–36.

Panagopoulos, C. and S. E. Ha. 2015. "Billboards and Turnout: A Randomized Field Experiment". *Journal of Political Marketing*. 14(4): 391–404.

Papas, R. K., H. L. Logan, and S. L. Tomar. 2004. "Effectiveness of a Community-based Oral Cancer Awareness Campaign (United States)". *Cancer Causes & Control*. 15(2): 121–31.

Schein, A., D. M. Blei, and D. P. Green. 2021. "Assessing the Effects of Friend-to-Friend Texting on Turnout in the 2020 US Presidential Election".

Shaw, C. 2014. *The Campaign Manager: Running and Winning Local Elections*. Westview Press.

Taylor, C. R., G. R. Franke, and H.-K. Bang. 2006. "Use and Effectiveness of Billboards: Perspectives from Selective-Perception Theory and Retail-gravity Models". *Journal of Advertising*. 35(4): 21–34.

Teresi, H. and M. R. Michelson. 2015. "Wired to Mobilize: The Effect of Social Networking Messages on Voter Turnout". *The Social Science Journal*. 52(2): 195–204.

Woodside, A. G. 1990. "Outdoor Advertising as Experiments". *Journal of the Academy of Marketing Science*. 18(3): 229–37.

Yancey, A. K., B. L. Cole, R. Brown, J. D. Williams, A. M. Y. Hillier, R. S. Kline, M. Ashe, S. A. Grier, D. Backman, and W. J. McCarthy. 2009. "A Cross-sectional Prevalence Study of Ethnically Targeted and General Audience Outdoor Obesity-related Advertising". *The Milbank Quarterly*. 87(1): 155–84.

Zekiri, J. 2019. "The Impact Of Billboard Advertising On Consumer Buying Behaviour". *Ecoforum Journal*. 8(2).