Blocks as geographic discontinuities: The effect of polling place assignment on voting

Sabina Tomkins[1], Keniel Yao[2], Johann Gaebler[2], Tobias Konitzer[3], David Rothschild[3,4], Marc Meredith[5], and Sharad Goel[6]

[1] *University of Michigan, Email: stomkins@umich.edu*
[2] *Stanford University*
[3] *PredictWise*
[4] *Microsoft Research*
[5] *University of Pennsylvania*
[6] *Harvard University*

Abstract

A potential voter must incur a number of costs in order to successfully cast an in-person ballot, including the costs associated with identifying and traveling to a polling place. In order to investigate how these costs affect voter turnout, we introduce two quasi-experimental designs that can be used to study how the political participation of registered voters is affected by differences in the relative distance that registrants must travel to their assigned Election Day polling place and whether their polling place remains at the same location as in a previous election. Our designs make comparisons of registrants who live on the same residential block, but are assigned to vote at different polling places. We find that living farther from a polling place and being assigned to a new polling place reduce in-person Election Day voting, but that registrants largely offset for this by casting more early in-person and mail ballots.

Keywords: Election administration, voting modalities, geographic discontinuity.

Political Analysis (2022)

DOI: 10.1017/pan.xxxx.xx

Corresponding author
Sabina Tomkins

Edited by
John Doe

1 Introduction

It has long been accepted by political scientists that the choice by a potential voter to vote or abstain from voting is based on an assessment of whether the benefits from voting are higher than the costs (Riker and Ordeshook 1968). The opportunity cost of the time spent casting a ballot is one of the most important costs that a potential voter must incur in order to vote (Downs 1957). There are concerns that even small increases in expectations about the time it takes to vote could reduce turnout, particularly given that many potential voters receive limited expected benefits from voting (Aldrich 1993).

Potential voters are likely to consider the expected search costs and transportation costs associated with voting when deciding whether to vote and which vote mode to use (Brady and McNulty 2011). Search costs refer to the cost of identifying where a polling place is located and how to get there, and are thought to decrease when a potential voter repeatedly votes at the same polling place. Transportation costs refer to the cost of traveling to a polling place, and will increase, typically, as a polling place moves further from a potential voter's residence.

Political science research shows that increases in search costs reduce the number of ballots voters cast in-person on Election Day. Brady and McNulty (2011) find that potential voters were two percentage points less likely to vote in-person on Election Day when they were assigned to vote at a new polling place that was located equally far from their residence as their old polling place. Two percentage points represents the median estimated reduction in in-person voting on Election Day from a polling place change in existing work, with McNulty, Dowling, and Ariotti (2009) and Amos, Smith, and Ste. Claire (2017) finding more than a two-percentage point decline, and Yoder (2018) and Clinton *et al.* (2021) finding less.

Political science research also shows that increases in transportation costs reduce the number of ballots cast in-person on Election Day. Most of the studies referenced in the previous paragraph find a greater reduction in the likelihood of voting in-person on Election Day as the distance between the new polling location and the potential voter's residence increases. Similarly, Cantoni (2020) shows that potential voters who live in the same neighborhood are less likely to vote when the polling place that they are assigned to vote at on Election Day is further from their residence.

Existing research differs in whether increasing the search and transportation costs associated with Election Day voting primarily causes potential voters to *abstain* or *substitute* to early in-person voting or mail balloting. Clinton *et al.* (2021) show that most potential voters dissuaded from voting in-person on Election Day by increases in search and transportation costs switched to early in-person voting. In contrast, Brady and McNulty (2011) find that about 60 percent of the potential voters who were dissuaded from voting in-person on Election Day because of higher search costs abstained, with the other 40 percent of shifting to mail ballots. Likewise, Amos, Smith, and Ste. Claire (2017) find that about 60 percent of the potential voters who were dissuaded from voting in-person on Election Day because of higher search costs abstained, with the other 40 percent of shifting to early in-person voting or mail balloting.

Synthesizing the existing evidence is challenging because studies vary both in the context that is being studied and the specific design being applied to conduct the study. States vary in the ease with which voters can substitute into using mail ballots or early in-person voting as the cost of Election Day voting increases. Some designs focus on the impact of transportation costs among those who are experiencing an increase in search costs, while others focus only on cases in which there are no changes in polling locations. Designs also differ in whether they focus exclusively on polling place changes resulting from the consolidation of polling places, or consider the larger set of polling place changes caused by consolidation, expansion, and movement of polling places. These differences make it hard to infer how the results of a given study may generalize to alternative contexts in which substitutes are more or less accessible or a different mechanism generates variability in voting costs.

We introduce two new quasi-experimental designs that can be applied to study the consequences of increased search and transit costs on potential voters in a wide variety of contexts. In brief, our designs leverage cases in which registrants who live on the same block of a street are assigned to vote at different polling places on Election Day. A common reason why this occurs is that a precinct boundary runs down the middle of a block so that registrants who live on the odd-numbered and even-numbered sides of the block are assigned to vote at a different polling places. This allows us to compare the turnout choices of registrants who by construction are similarly situated, but are traveling to polling places that potentially differ either in terms of their proximity to the block or their stability over elections.

It is easy to incorporate data from many types of jurisdictions using our designs, allowing us to generate the most comprehensive analysis to date of registrants' ability to offset higher costs of Election Day voting with substitutes. The key variables that we need to observe to identify potential variation in the cost of in-person voting are residential address and precinct, which are observable in any voter registration database. Moreover, our approach does not require states to provide any additional information beyond the mapping between precincts and the addresses of the precincts' polling locations. Importantly, this means we are not limited to only studying states that make it easy to access the mostly non-existent data files of precinct boundaries. Our primary analysis focuses on data from ten states in which we were able to: (1) collect information of polling place addresses in the 2012 and 2016 presidential elections; (2) link these data with voter registration records; and (3) discern from the voter registration record whether the 2016 ballot was cast on Election Day or using a substitute vote method (i.e., mail-ballot or early in-person ballot). We also perform some additional analyses for the six of these ten states in which there were sufficient observations to estimate the effect of a registrant being assigned to vote on Election Day at a different polling location in 2016 than in 2012.

Our results show that the ability to substitute into other modes of voting is key to helping mitigate consequences of increased search and transit costs on Election Day voting. We find that a registrant who lives further from their polling places is approximately 1.5 percentage points (p.p) less likely to vote at this polling place than a similarly situated registrant who lives closer, but offsets for this with their increased use of mail balloting or early in-person voting. Likewise, a registrant assigned to a different polling place in 2016 than in 2012 is about 1.3 p.p. less likely to vote in-person on Election Day than a similarly situated registrant who was assigned to the same polling place in both elections, but partially compensates for this by substituting into mail balloting or early in-person voting. These findings highlight the importance of making alternative voting methods available for potential voters who find it difficult to vote at their polling place on Election Day.

2 Data

Our analysis uses state-supplied polling place files, which provide identifiers and addresses for polling places, and a national voter file, which provides registrant characteristics, turnout information, and voting place identifiers. Together these data allow us to infer which polling place registrants were assigned to vote at on Election Day and to compute the distance from their residences to this polling place.[1] We then select blocks comprised of similar registrants who are assigned to vote at different polling places and use variation in distance between the block and the polling places, as well as the stability of the location of polling places between two elections, to estimate the effects of relative distance and changes to polling places (henceforth shocks) on turnout decisions. The remainder of this section provides a sketch of the process of cleaning and combining our primary data sources, with more details provided in Section A.4 of the Supplemental Appendix.

To collect data on polling place locations, we filed public-records requests to state-election officials in all fifty states. The full account of these requests is included in Section A.3 of the Supplemental Appendix. We received records from 18 states which met the basic criterion that they contain address and voting jurisdiction descriptors for both 2012 and 2016. A number of states, including Texas, reported that there was no statewide aggregation of polling place locations. A subset of the 254 counties in Texas also responded to our public record requests with data that met the basic criterion. Table 1 shows a stylized example of the information these data contain.

We merge information on polling place locations to registration records in snapshots of TargetSmart's national voter file from November 2012 and November 2016. These data include a registrant's voting jurisdiction (i.e., precinct), registration address, and vote history.[2] We restrict our analysis to the subset of states in which the vote history data differentiates if a vote was cast in-person on Election Day or using a substitute vote method. While we would like to separate substitute votes that are cast early in-person and by mail, this is not something that most of the states in our analysis differentiate in their vote history.

Our design requires us to find blocks of registrants in which some registrants on the block are assigned to vote at one polling location on Election Day, while other registrants on the same block are assigned to vote at a different polling location on Election Day. We create a block identifier for each address of registration that consists of all but the final two digits of the street number (e.g. 200 would be encoded as 2, and 2100 would be encoded as 21), the street name, the street type, and the city and state of a voter's residential address. For example, the two registrants in Table 2, residing at 123 Main St. and 125 Main St. in Milwaukee, Wisconsin, respectively, share a block identifier. Table 2 shows a stylized example of these data after these block identifiers are generated.

To infer polling-place assignments we match registrants to polling places with voting-jurisdiction metadata, such as precinct name and city. Consider the three example voters in Table 2 and the two example polling places in Table 1. Voter-1 and Voter-2 would be assigned to PP-1 while Voter-3 would be assigned to PP-2. We detail in Section A.5 of the Supplemental Appendix how we calculate distance between a registrant's residential address and the assigned polling place once we construct these data.

We apply filters to our data after matching registrants to polling places to ensure that registrants in the same block are similar except potentially for the location of their polling place in our baseline analysis. First, our analysis focuses on registrants with the same registration address in the 2012

1. Data can be found here: Tomkins *et al.* 2022

2. The TargetSmart's national voter file supplements the public information available in each state's voter file with additional information they either collected or modeled. While most of the data that we use from the TargetSmart's national voter file could be observed in the individual state's vote files, we do make use of a few variables that are not. Most importantly, we rely on TargetSmart's geocodes of registrants' registration address. We also use information that TargetSmart collected or modeled on a registrant's party, race, and home value when assessing the similarity of registrants assigned to different types of polling locations.

Place ID	Address	City	State	Precinct
PP-1	200 Main St	Milwaukee	WI	Cherry School 1
PP-2	1000 Third St	Milwaukee	WI	Apple School 1

Table 1. Example polling place file.

Voter ID	Address	City	State	Block ID	Precinct
Voter-1	123 Main St	Milwaukee	WI	1-Main-St-Milwaukee-WI	Cherry School 1
Voter-2	125 Main St	Milwaukee	WI	1-Main-St-Milwaukee-WI	Cherry School 1
Voter-3	2000 Third St	Milwaukee	WI	20-Third-St-Milwaukee-WI	Apple School 1

Table 2. Example voter records.

and 2016 snapshots to ensure changes in the location of polling places result from changes in polling place assignment rather than registrants moving. Second, we restrict our sample to blocks in which registrants are assigned to two distinct polling locations. We also require that at least two registrants were assigned to both of these polling locations. Third, we discard blocks in which there are two registrants on that block that live more than 0.3 miles apart. In the Supplemental Appendix, we show that the results of our relative distance analysis are very similar if we eliminate the filter that registrants must live in same registration address in the 2012 and 2016 and the results of both our relative distance and shocks analyses are robust to smaller or larger filters on how far away two registrants on that block can be.

3 Methods: Measuring the effects of relative distance and shock

Polling-place assignments dictate many aspects of registrants' voting experiences, including how long it takes to cast a ballot and how far one must travel to do so. We propose identification strategies that exploit the quasi-random assignment of registrants living on the same block to different polling places to estimate two causal effects.

We refer to the first effect as the *effect of relative distance*. Here we consider two simplified notions of distance when registrants on a block are assigned to two polling places: farther and nearer. All registrants on a block assigned to the nearer polling place are said to experience a nearer distance, while those on the face with a greater average distance are said to experience a farther distance. The *effect of relative distance* is the difference in the likelihood that those registrants who experience a farther distance vote using a given method relative to those registrants who experience a nearer distance.

We refer to the second effect as the *effect of shock*. A shock occurs when a registrant, living at the same residence, is reassigned to a new polling place in between elections. The *effect of shock* is the difference in the likelihood that those registrants who experience a shock vote using a given method relative to comparable registrants who were assigned to vote at the same polling place in both elections. Additionally, we require that one of two conditions holds: either that all registrants were assigned to the same polling place in 2012, or, if registrants were assigned to two different places that the average distance to each place be similar (that the difference in the average distance to each polling place be less than 0.25 miles).

We are interested in the effect of relative distance and the effect of shock on turnout, as well as voting using specific methods. Thus, we examine the effect of both on all voting (any method), voting in person on Election Day (in person), and voting by a substitute method such as casting a mail-in or absentee ballot, or voting early in person (substitution).

For each of the voting behaviors we examine, let

$$Y_i^m = \begin{cases} 1, & \text{if registrants } i \text{ cast a vote by method } m, \\ 0, & \text{otherwise,} \end{cases}$$

where $m \in \{\text{in person, substitution, any method}\}$.

The linear regression specified in Equation 1 is used to estimate the corresponding effect of relative distance and effect of shock. In Equation 1, $\mathcal{T}_b$ is the set of all registrants on block b assigned to treatment, $\mathcal{C}_b$ is the set of all registrants on block b assigned to control, $treatment_i = \mathbb{I}_{i \in \mathcal{T}_b}$ is an indicator of whether a registrant is assigned to treatment or control, and γ_b is a block fixed effect. Additionally, $h(i)$ refers to the household of registrant i. When estimating Equation 1, standard errors are clustered by household to account for autocorrelation in the unmodeled determinants of voting behavior by registrants who reside in the same household.

$$Y_i^m = \gamma_b + \theta treatment_i + \epsilon_{i,h(i)} \tag{1}$$

3.1 Illustrating our identification strategies

To estimate the effect of relative distance we identify all blocks which lie on the boundaries of two voting jurisdictions. For example, we show a block where one face (shown with a solid line) is assigned to a different polling place than the face across the street (shown with a dashed line) in the top panel of Figure 1. To compare the turnout behavior of the registrants who live farther from their polling place relative to those on the same block who live closer, we require that a block contain registrants assigned to two different polling places. Additionally, we require that these assignments remain unchanged between 2012 and 2016 as this allows us to isolate the effect of relative distance independent of effect of shock. For such blocks, registrants assigned to the polling place with the lesser average distance to its registrants are assigned to control, and registrants assigned to the place with the greater average distance are assigned to treatment. This approach is similar to one used by Middleton and Green (2008), which used variation in canvassing activity along the boundary of a precinct to estimate the effect of canvassing on turnout. The set of all relevant blocks defined by this identification strategy spans ten states (Hawaii, Iowa, Indiana, Maryland, North Carolina, Pennsylvania, Rhode Island, Texas, Utah and Wisconsin) and a total of 252,428 registrants.

Our identification strategy for the effect of shock is illustrated in Figure 1. We locate all blocks which lie on the boundaries of two precincts, such that a single face of this block experiences a polling-place shock between 2012 and 2016 while the other does not. Figure 1 highlights two block faces outlined with dashed and solid borders. All registrants who reside on the face enclosed with the dashed line experience a shock (and are assigned to treatment) and all those residing on the side enclosed with the solid line do not (and are assigned to control).

Note that a shock may arise for multiple reasons. First, a reduction in the number of polling places, which is sometimes referred to as a consolidation, causes some registrants to be assigned to a new polling place that, on-average, will be located further away from their residence. Second, an increase in the number of polling places causes some registrants to be assigned to a new polling place that, on-average, will be located closer to their residence. Finally, a polling place being moved causes some registrants to be assigned to a new polling place without any clear expectation about how the change will affect the average distance between the polling location and the registrants' residence. Here, we do not differentiate between types of shocks; any block where one block face experiences a change and the other does not is included in our analysis regardless of whether the change is a result of a consolidation, addition or a movement. This contrasts our approach with some previous studies which focus specifically on the effects of consolidations (Brady and McNulty 2011; Amos, Smith, and Ste. Claire 2017).

We use this identification strategy to select relevant blocks from the dataset we described the construction of in Section 2. In addition to the criteria detailed in Section A.4 of the Supplemental Appendix, we require that each block face on the same block be a similar distance from its assigned polling place in 2012 in the blocks used to estimate the effect of shock. Specifically, we require that the difference between the average distance to the polling place on the two block faces be no more than 0.25 miles different. In total, the dataset used to estimate the effect of shock spans six states (Iowa, Indiana, Maryland, North Carolina, Pennsylvania and Wisconsin) and includes 47,321 registrants.

These identification strategies are examples of geographic discontinuity designs, which have been recently popularized in political science by L. J. Keele and Titiunik (2015) and L. Keele and Titiunik (2016). Focusing only on comparisons within blocks will cause us to ignore many potential comparisons of registrants who live in close proximity to one another, but are assigned to vote at different polling places. The benefit is that we expect there to be fewer differences in the underlying propensity to vote among registrants on the same block than registrants living on different blocks that are located in close proximity. And even when there are differences, there is no reason to expect that registrants with a higher propensity to vote should systematically end up closer or further from their polling place than those with a lower propensity to vote. Thus, focusing on registrants residing on the same block allows us to implement a geographic discontinuity design even when we know little about the specific registrants, and when there are too many boundaries to apply a method like that proposed by L. J. Keele and Titiunik that empirically investigates whether housing prices are comparable on either side of a given boundary to determine whether we expect the underlying propensity to vote to be similar on either side of the boundary.

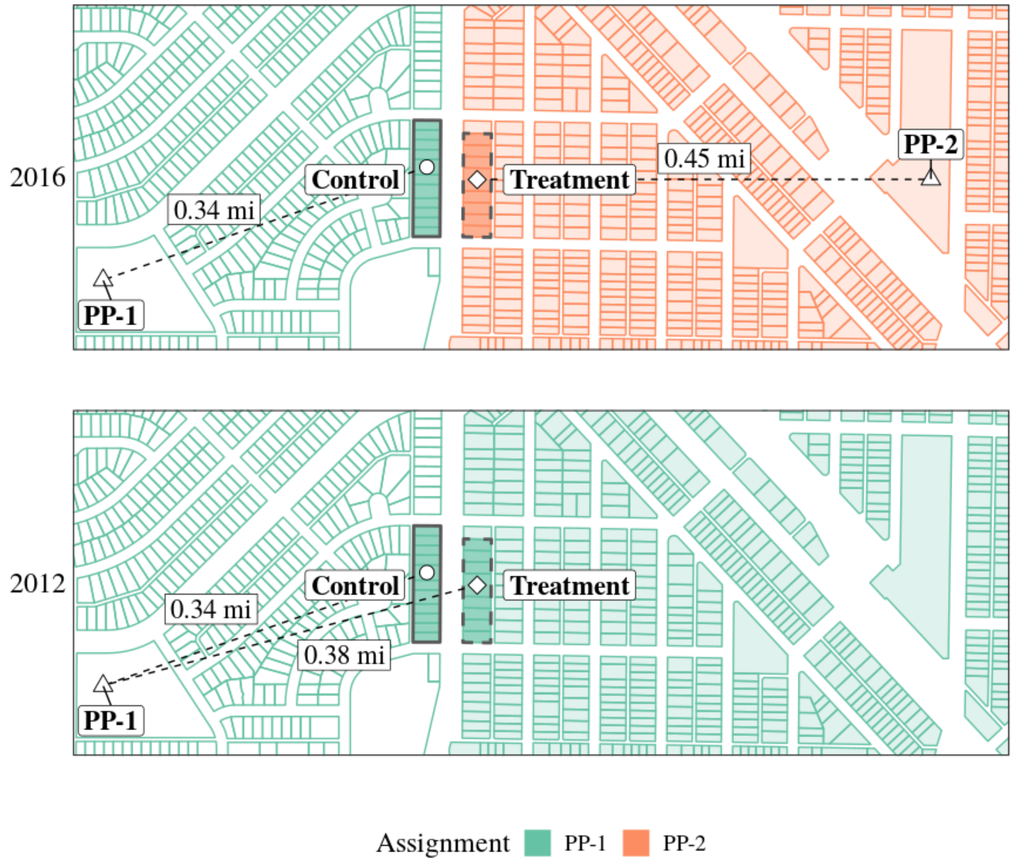


Figure 1. This map illustrates the block-randomization identification strategy for the effect of shock with two precincts in Milwaukee, WI across 2012 and 2016. Each precinct's color corresponds to its polling place assignment. In 2012 all registrants in both precincts are assigned to PP-1. In 2016 all registrants of the right-most precinct experience a shock as their polling place assignment changes from PP-1 to PP-2. To identify eligible blocks of voters for our analysis, we identify blocks on the boundaries of the two precincts; registrants of the block face that experiences a shock (dashed outline) are assigned to treatment while registrants of the block face that does not experience a shock (solid outline) are assigned to control.

3.2 Threats to internal validity

Our identification strategy assumes that registrants who live on different sides of a block would use voting methods at the same rate if not for differences in the polling places at which registrants on each side of a block were assigned to vote on Election Day. The assumption may be not be proper, for example, if there is sorting that results in more politically active registrants being systematically more likely to reside on the of side of a block that gets assigned to vote at a closer or more stable polling location. In this context, we doubt that the mechanism causing sorting would be registrants thinking about differences in polling locations when deciding on which side of the street to reside. Rather, we speculate that sorting would be more likely to be caused by politically active registrants potentially being better able to exert influence over where polling places are located.

A common way to assess the assumption that two groups would behave the same absent some treatment is to look at whether there any systematic differences in their observable characteristics. The figures that we present in Section A.1 of the Supplemental Appendix examine how the characteristics of registrants assigned to different types of polling places differ. We first look at age, gender, modeled partisanship, and modeled race, as well as indicators representing that the total number of registrants on that side of the block fell into a certain range and find that all absolute differences are less than 2 percentage points and that the majority are less than 1 percentage point. For a subset of registrants for which we observe sale prices of homes, we follow L. J. Keele and Titiunik (2015) and examine whether the housing prices are comparable on the side of blocks that are closer and further from their polling place. We find that housing prices are similar for each side of the block. The lack of a clear difference in housing prices seems most important given that we might expect wealthier registrants to be able to exert more influence over where a polling place is located than less wealthy registrants.

Similar patterns are found when we conduct the same analysis except defining those on the side of a block experiencing a polling place change as the treatment group and those on the side of a block keeping the same polling place as the control group. Unlike with our relative distance analysis, we also expect to observe similar voting patterns in the treatment and control groups in the previous election before the shock was realized. This is indeed the case, as we find similar historical voting patterns in both groups (see Section A.1 of the Supplemental Appendix).

Post-treatment bias is another threat to internal validity that could also be present given our identification strategy. We focus on the turnout behavior of people who are registered to vote. If people are not registered because their polling place is located too far from their residence or because their polling place moved to a new location, this could bias our understanding of effect effect of relative distance and the effect of shock on voting behavior (Nyhan, Skovron, and Titiunik 2017). The effect of the post-treatment bias on our estimates could either be positive or negative depending on the frequency with which registrants would have voted, and the vote mode they would have used if voting, had they remained on the registration rolls. We are more concerned about the potential consequences of post-treatment bias when estimating the effect of relative distance than the effect of shock because shocks are unlikely to cause registrations to be immediately removed from the rolls given the registration removal process specified by the National Voter Registration Act of 1993.

To assess whether there is any evidence suggesting treatments affect registration status, we examine whether there are any differences in the likelihood that treated and control registrants remain registered to vote in 2016 at the same address they were registered to vote at in 2012. Those who resided further from the polling place in 2012 were 0.14 ± 0.27 p.p. less likely to have such a voter registration in 2016 than those who resided closer. Turning to the shock analysis, we found that 2012 registrants who experienced a polling place change were 0.40 ± 0.80 p.p. less likely to be registered as the same address in 2016 than 2012 registrants whose polling place remained the same.[3] While certainly not conclusive, we are not observing evidence of substantial differential attrition by treatment status in the data.

3.3 Threats to external validity

One potential threat to the external validity of our analysis is that it is pooling together relative distance differences of varying magnitudes to estimate the effect of relative distance. This limits the ability to use our analysis to learn about the potential consequences of a given change in relative distance on voting behavior. One reason why we do this is that we do not think that a causal effect of a one-unit change in relative distance is a well-defined concept. The amount of time that it takes someone to travel to their polling place is the underlying causal variable that we think affects the choices of whether to vote and which vote method to use. Travel time is clearly related to, but not the same, as the distance someone travels from their residence to the polls. How distance translates into travel time will be affected by the mode of transit someone uses to get to the polls (e.g., walking, bus, car) and the speed at which someone can use that mode of transit to cover that

3. We report 95% confidence intervals throughout in the paper.

distance (Bhatti 2012; de Benedictis-Kessner and Palmer 2020).

Our current analysis requires an assumption that when similar people use the same modes to travel on similar roads, it will take longer, on average, to travel a longer distance than a shorter distance. We then use our design to identify cases in which we think people should be similar in terms of their access to cars and the types of roads that they are traveling on because they live on the same block. Thus, we feel comfortable asserting that the average cost of Election Day voting is higher for registrants who have to travel a further distance to their polling place than other registrants who live on the same block who travel a shorter distance. And by constraining them to live on the same block, we also enforce that the cost of accessing early in-person voting (if it exists) should be similar in expectation.

An analysis that incorporates cardinal information about distance requires additional assumptions about how differences in the relative distances across blocks translate into differences in travel times in order to estimate a causal effect. We do not think that the assumptions necessary to estimate a causal effect of this form are likely to hold. For example, consider a block on a city's main street on which a registrant on block face A travels 0.4 miles further to their polling place than a registrant on block face B. And compare that to a block on a rural road on which a registrant on block face C travels 1.2 miles further than a registrant on block face D. It could be that 0.4 miles means more to the A vs. B residents than the 1.2 miles means to C vs. D residents, especially if registrants on block faces A and B are more likely to walk to the polls than registrants on block face C and D. In the conclusion, we discuss how future work might expand upon the work that we do here to incorporate more information about the degree of differences in relative distance into the analysis.

A related concern about external validity is that we are estimating the effect of relative distance and effect of shock for the set of blocks in our sample for which we observe different registrants assigned to vote at different polling places. Not only is there nothing that guarantees that these blocks are representative of the broader population of blocks, but the types of blocks we need to implement our identification strategy – those with registrants assigned to vote at multiple polling places – may be particularly likely to arise in densely populated urban settings. Thus, if there is heterogeneity in the effect of relative distance and effect of shock in urban and rural contexts, then our conclusions about the effect of relative distance and effect of shock are likely to over represent the effects in urban settings and under represent the effects in rural settings.

4 Results

We first report the results of the effect of relative distance. The left panel of Figure 2 shows that registrants that live relatively further from their assigned polling places are generally less likely to vote at their assigned polling place on Election Day than registrants that live relatively closer. The solid orange bars show that in eight of the tens states we study, registrants who live relatively further from their assigned polling place were less likely to vote at that polling place on Election Day than registrants who live relatively closer. The error bars represent 95 percent confidence intervals for each state as well as our aggregate estimate. Pooling over all ten states, a registrant that lives relatively further from their assigned polling place is 1.5 (±0.40) p.p. less likely to vote at their assigned polling place on Election Day than a registrant that lives relatively closer.

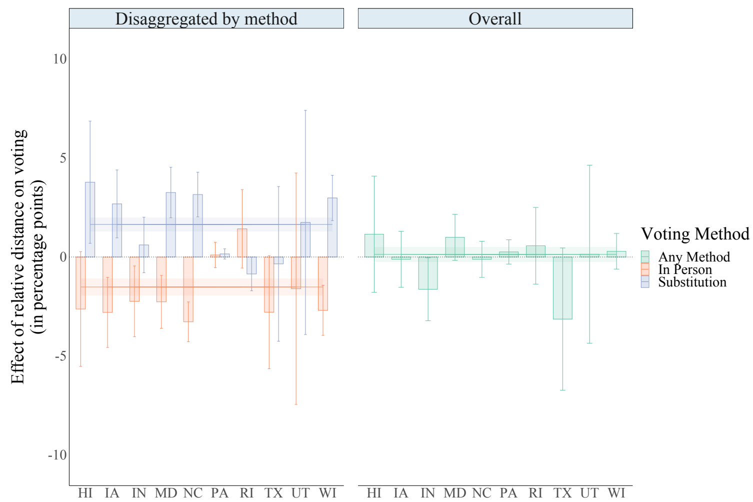


Figure 2. Relative distance does not affect the likelihood of voting, but does affect the method used to vote.

The remainder of Figure 2 shows that registrants that live relatively further from their assigned polling location compensate for a lower rate of voting at their assigned polling place with higher rates of early in-person voting or mail-in balloting. The solid purple bars show that in eight of the tens states we study, registrants who live relatively further from their assigned polling place were more likely to use early in-person voting or mail-in balloting than registrants who live relatively closer. Pooling over all ten states, a registrant that lives relatively further from their assigned polling place is 1.6 (±0.58) p.p. more likely to use early in-person voting or mail-in balloting than a registrant that lives relatively further. The right panel shows that there is almost no difference in the turnout rate of registrants who live relatively further or closer from their assigned polling place.

Figure 3 helps to contextualize how much further registrants who live relatively further from their assigned polling location have to travel in order to vote in-person on Election Day. Figure 3 shows the average additional distance that these registrants must travel as compared to registrants who live relatively closer in a random sample of 1,000 blocks included in our analysis. Even though 85% of all distance differences are less than one mile, this is still sufficient to generate the effect of relative distance on both voting in person and by substitution observed in Figure 2. Figure 3 also demonstrates that the additional distance that registrants that live relatively further from their assigned polling location must travel to that polling location increases as a function of the distance that registrants that live relatively closer to their assigned polling location must travel to that polling location.

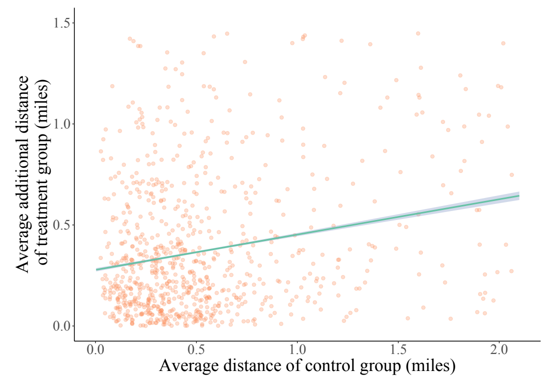


Figure 3. Average distance to the polling place among registrants who live relatively closer and the average additional distance registrants who live relatively further must travel in a sample of 1,000 blocks in our *effect of relative distance* analysis.

Figure 2 shows that substitution methods appear to offer an alternative to registrants when the cost of voting on Election Day increases. However, access to substitutions varies between states and even within states. For example, some states allow for both no-excuse mail ballots and early in-person voting (e.g., North Carolina), while others allow no-excuse absentee voting only for certain voters (e.g., Indiana). Thus we inspect how voting by substitution varies with a state's openness to substitution adoption. As a proxy for a state's openness to substitution adoption, we use the percentage of voters that cast a mail or early in-person ballot in the 2012 presidential election. Consistent with our expectations, Figure 4 shows that registrants who live relatively further from their assigned polling location are least likely to respond with substitution in states with the lowest level of adoption. However, we cannot rule out the possibility that there may be other state-specific factors that affect substitution patterns and the use of mail or early in-person ballots.

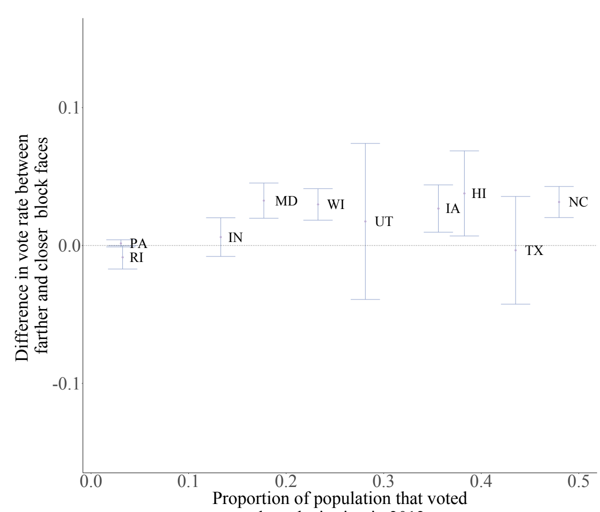


Figure 4. Registrants who live relatively further from their assigned polling location are least likely to respond with substitution in states with the lowest usage of mail or early in-person ballots.

Next, we address the question of how a shock affects Election Day voting, voting by substitution and voting by any method. In Figure 5 we consider the data produced by the effect of shock, where one block face experiences a polling place assignment shock in 2016. It shows that shock creates a shift in how voting occurs, reducing voting in person by 1.3 (±1.00) p.p. and increasing voting by substitution by 0.76 (±0.71) p.p. Overall, we see that as with effect of relative distance, effect of shock is moderated by registrants voting by substitute modalities rather than in person.

Figure 5. While shocks reduce Election Day voting, more that half of this reduction is compensated for by increased voting by substitution.

One complication with examining shocks is that registrants experiencing shocks often also experience changes in the relative distance between their residence and their polling place. Figure 6 shows that some shocks cause registrants to reside closer to their polling place, while other shocks cause registrants to reside further. One reason for this heterogeneity is that the shocks in our data result from a combination of counties adding polling places, consolidating polling places, and moving polling places. Each of these events might influence the cost of voting differently. Consolidations may not only increase the average distance registrants travel to their polling place, but also increase the cost of voting via longer lines. Conversely, additions may make Election Day voting more convenient in other ways beyond reducing the average distance registrants travel to their polling place. Table A.1 in the Supplemental Appendix shows that more registrants who experience a shock live in a county that added polling places than a county that subtracted polling places, suggesting that additions may be more likely to generate the shocks in our data than subtractions.

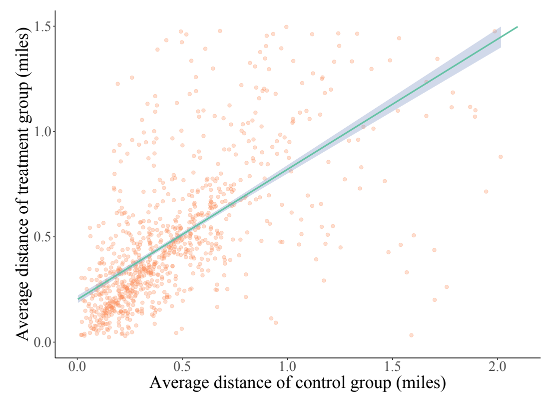


Figure 6. Average distance to the new polling place among registrants who experience a polling place change in a sample of 1,000 blocks in our *effect of shock* analysis.

Figure 7 examines whether the effect of shock on Election Day voting is more pronounced when registrants must travel greater distances than they were previously required. It inspects the difference in the voting rate between the block side which experienced a shock and the side that did not as a function of the difference in the distance of the block from the new polling place and the distance of the block from the old polling place. The x-axis varies which blocks are included when estimating the effect of shock based on this difference in distance. For example, the estimate reported for "> 0.5 miles" only includes blocks that experienced a shock which resulted in the new polling place being 0.5 miles or more further from the block than the old polling place.

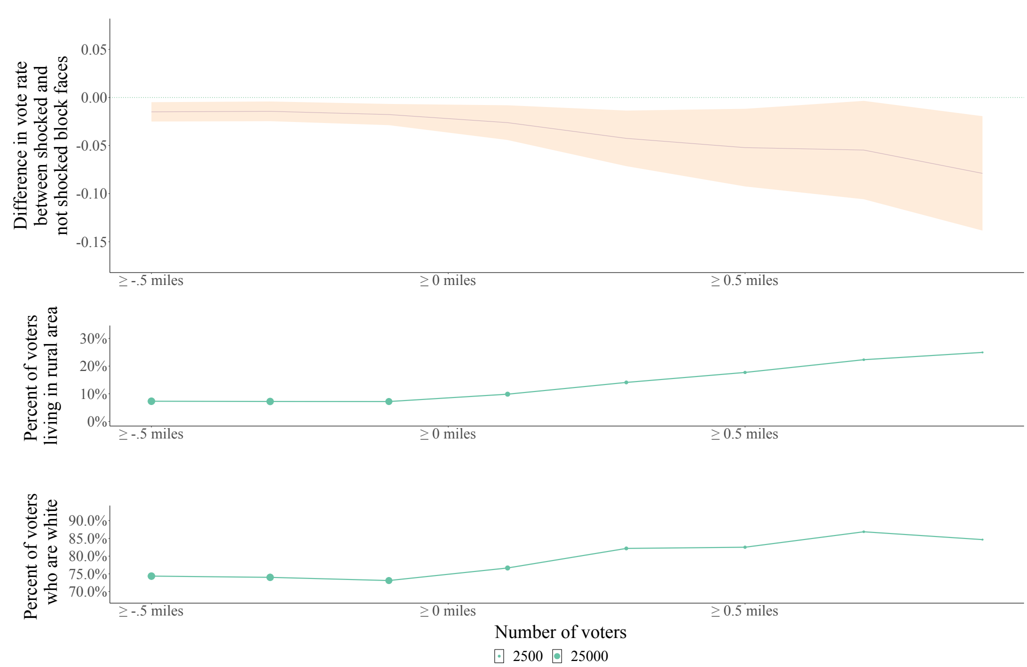


Figure 7. Shocks cause a greater reduction in Election Day voting when the sample of blocks is restricted to those where difference in distance to the polling place is greater among those who did and did not experience a shock.

The top panel of Figure 7 shows that registrants that live further from their new polling place than their old polling place (i.e. x-axis equal 0 miles) are about -2.0 ± 1.3 p.p. less likely to vote on Election Day than people who live on the same block who did not experience a polling place change. While it cannot be observed directly on the graph, the corresponding estimates for those registrants whose new polling place is closer to their residence than their old polling place is -0.5 ± 1.5 p.p. Why might Election Day voting decline, albeit by a slight amount, among registrants that experience a reduction in transportation costs? We think the most likely explanation has to do with election administration procedures. Often times registrants who are experiencing a polling place change are sent notices from election officials, either by statue or custom, that inform them of the change. While sometimes these notices explicitly include information about substitute voting methods that are available, we expect some recipients of such a notice to be induced to research substitute voting methods even when they are not referenced on the notice. Another explanation could be that registrants abstain from voting on Election Day because they are angry about the change, although this is probably less to be true when registrants are moved to a polling location that is closer to their residence. Finally, registrants who are unaware of a polling place change may show up to vote at the incorrect polling place, which could cause a decline in Election Day voting if these registrants do not subsequently move onto their new polling place to cast a ballot.

The top panel of Figure 7 also shows a clear trend that larger increases in distance that registrants must travel to their new polling place relative to their old polling place associate with larger reductions in Election Day voting. Similarly, Figure A.8 in the Supplemental Appendix shows that the effect of relative distance on Election Day voting is larger the more we restrict the sample to blocks where the difference in distance between the further and closer polling place is larger. However, the middle and bottom panels of Figure 7 demonstrate why this cannot necessarily be interpreted as the effect of distance. The middle and bottom panels show that the underlying population used to construct each estimate varies, with shocks causing larger differences in distance to the polling place in shocked blocks in rural areas. Moreover, shocks that cause registrants to reside substantially further from their polling place are disproportionately likely to be caused by consolidations. We leave it to future work to isolate the simultaneous effect of a unit increase in distance on registrants experiencing polling place assignment shocks.

We include a number of additional plots in the Supplemental Appendix to establish the robustness of our findings to alternative assumptions. Our baseline analysis restricts our sample of blocks to those in which all registrants are within .3 miles of each other. Our choice of a .3 mile cutoff in our baseline analysis is based on visual inspection of the data and observing that residences on what are commonly understood to comprise a block rarely are more than .3 miles apart from one another. But occasionally addresses are assigned in a different way such that two residences that are located quite far apart are also on same block. Figure A.9 in the Supplemental Appendix shows we reach nearly identical conclusions when we apply a cutoff of .1 miles or .5 miles instead of .3 miles. Our baseline analysis also only considers registrants who were registered at the same address of registration since 2012. While such a restriction is necessary to identify registrants who experienced a polling place shock, we can estimate the effect of relative distance on the full set of 2016 registrants. Figure A.10 in the Supplemental Appendix demonstrates nearly identical results when we remove the restriction that registrants continued to reside at the same address of registration since 2012 when estimating the effect of relative distance.

5 Discussion

This paper presents the most comprehensive analysis to date of the effect of relative distance and the effect of shock on voter turnout. It is able to do so because we develop two quasi-experimental designs that only require knowledge of registrants' residential addresses and polling place addresses, allowing us to utilize data from multiple states. We find clear evidence that registrants who live further from their polling place and who experience a polling place change are less likely to vote at their polling place on Election Day. The reduction in Election Day voting caused by having a longer trip to the polls appears to be completely offset by increases in early in-person or by mail-in voting. Likewise, our point estimates suggest that a majority of the reduction in Election Day voting caused by shocks are offset by increases in early in-person or by mail-in voting, and that overall registrants who experience shocks are similarly likely to vote as those who live on the same block who do not.

The finding of minimal effects of shocks on overall turnout contrasts with many of the previous studies of shocks. We speculate that one reason why our findings differ from previous work is that most previous work focuses specifically on shocks resulting from the consolidation of polling places. In contrast, we find that in our data registrants are more likely to experience shocks in counties that are adding polling places than in counties that are reducing polling places, suggesting that many of the shocks in our data are not caused by consolidation. While all shocks impose search costs, shocks that result from consolidation are likely to also impose other additional costs on voting. Conversely, shocks that result from polling place exchanges, or especially additions, may reduce the cost of voting in other ways.

Our results also contrast with Cantoni (2020), which is the only other paper that studies the effect of relative distance independent of shocks. While Cantoni finds a substantial decrease in overall turnout among registrants located further from their polling place relative to similarly situated registrants that are located closer, we find registrants who need to travel further to their polling place compensate for their reduced Election Day voting with greater early in-person and mail voting. There are a number of potential explanations for why we reach different conclusions. First, the ability for registrants to switch to early in-person or mail balloting likely depends, in part, on the accessibility of these alternative voting methods. Second, registrants may not respond differently to small and large differences in relative distance. Most blocks in our sample were within a mile of their assigned polling place and the registrants who lived further from their polling place usually lived less than 0.5 miles further from their polling place than the registrants who lived closer. Finally, the effect of relative distance may depend on the characteristics of the specifics registrants and elections being studied. Cantoni finds, for example, the effect of relative distance is greater in areas where car ownership is lower, and is lower in the 2016 presidential election than in the other elections that he studied.

While our identification strategy can be applied broadly in settings where registrants are assigned to polling places, it is limited in some respects. First, it only allows us to estimate the effect of relative distance and the effect of shock on the turnout of registrants. If the experience of greater distance or shocks causes potential voters to not be registered, then our design may mismeasure the effect of these variables on turnout (Nyhan, Skovron, and Titiunik 2017). Moreover, our focus on registrants who lived at the same address for four years means our sample under-represents the effect of relative distance and the effect of shock on potential voters with less residential stability. We also require that a registrant's address can be reliably geocoded. Perhaps because rural addresses might be more difficult to reliably geocode and we require all registrants on a block to live within 0.3 miles of each other, registrants from rural areas appear to be underrepresented in our analysis.

Future work can build off the identification strategy we propose here to develop more complex measures of distance. Here, we simplify all measures of continuous distance into a simple notion of farther or closer. This distinction allows us to compare registrants across geographies in a way which we believe is statistically defensible. However, this means our estimated treatment effects bundle together many comparisons of the effect of relative distance that likely affect voting behavior in different ways. We think the most promising approach would follow Clinton *et al.* (2021) and directly translate differences in relative distances into expected differences in travel time. Doing so requires paying more attention to how people travel to their polling places and how this varies across different types of neighborhoods. One feature of the designs that we put forward that might be beneficial is that we can apriori identify the set of registrants for which it is important to construct good measures of expected travel time. Limiting the number of registrants that we need to consider may increase our capacity to generate accurate measures of expected travel time for these cases.

Our work is motivated by a desire to understand the magnitude of the costs of voting on Election Day. While we caution that these are limited to the population of registered voters, an optimistic result of our work is that these costs are readily offset by the ability of registrants to shift into substitute modalities. The provision of mail-in ballots and early voting venues can have important implications beyond the ability to vote, but perhaps on the experience of political engagement. We leave it to future work to explore not only the ability of various voting modalities to offset costs but the implications of this effect for all potential voters.

Supplementary Material

(This is dummy text) For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.xxxx.xx.

References

Aldrich, J. H. 1993. "Rational choice and turnout." *American Journal of Political Science* 37 (1): 246–278.

Amos, B., D. A. Smith, and C. Ste. Claire. 2017. "Reprecincting and voting behavior." *Political Behavior* 39 (1): 133–156.

Bhatti, Y. 2012. "Distance and voting: Evidence from Danish municipalities." *Scandinavian Political Studies* 35 (2): 141–158.

Brady, H. E., and J. E. McNulty. 2011. "Turning out to vote: The costs of finding and getting to the polling place." *The American Political Science Review* 105 (1): 115–134.

Cantoni, E. 2020. "A precinct too far: Turnout and voting costs." *American Economic Journal: Applied Economics* 12 (1): 61–85.

Clinton, J., N. Eubank, A. Fresh, and M. E. Shepherd. 2021. "Polling place changes and political participation: Evidence from North Carolina presidential elections, 2008-2016." *Political Science Research and Methods* 9:800–817.

de Benedictis-Kessner, J., and M. Palmer. 2020. "Driving turnout: The effect of car ownership on electoral participation." *HKS Faculty Research Working Paper Series RWP20-032.*

Downs, A. 1957. *An economic theory of democracy.* Harper / Row: New York, NY.

Keele, L., and R. Titiunik. 2016. "Natural experiments based on geography." *Political Science Research and Methods* 4 (1): 65–95.

Keele, L. J., and R. Titiunik. 2015. "Geographic boundaries as regression discontinuities." *Political Analysis* 23 (1): 127–155.

McNulty, J. E., C. M. Dowling, and M. H. Ariotti. 2009. "Driving saints to sin: How increasing the difficulty of voting dissuades even the most motivated voters." *Political Analysis* 17 (4): 435–455.

Middleton, J. A., and D. P. Green. 2008. "Do community-based voter mobilization campaigns work even in battleground states? Evaluating the effectiveness of MoveOn's 2004 outreach campaign." *Quarterly Journal of Political Science* 3 (1): 63–82.

Nyhan, B., C. Skovron, and R. Titiunik. 2017. "Differential registration bias in voter file data: A sensitivity analysis approach." *American Journal of Political Science* 61 (3): 744–760.

Riker, W. H., and P. C. Ordeshook. 1968. "A theory of the calculus of voting." *American Political Science Review* 62 (1): 25–42.

Tomkins, S., K. Yao, J. Gaebler, T. Konitzer, D. Rothschild, M. Meredith, and S. Goel. 2022. https://doi.org/10.7910/DVN/45GQNA. V. DRAFT VERSION. https://doi.org/10.7910/DVN/45GQNA.

Yoder, J. 2018. "How polling place changes reduce turnout: Evidence from administrative data in North Carolina." *Available at SSRN 3178184.*

6 Appendix

6.1 Additional Figures

Figure 6.1 shows how the demographics of registrants who live on the side of the block that is further from its polling place (i.e., treatment group) compares to the demographics of registrants who live on the side of the block that is closer to its polling place (i.e., control group). We inspect age, gender, modeled partisanship, modeled race, as well as the number of registrants on either side of the block. Each point is the difference between the proportion of registrants with a given characteristic in treatment and the proportion of registrants with this characteristic in control, divided by the proportion of registrants with a given characteristic in control. The size of each point is scaled by the number of registrants in the control group, where larger points represent groups with more registrants.

The difference between the proportion of registrants with a given characteristic in the treatment group relative to the control group is close to zero for the most common characteristics. For less common characteristics the differences can be greater. Here, we show the percent differences. However, note that all absolute differences are less than 2 percentage points and that the majority are less than one percentage point.

We also observe information on the sale prices of homes for about 14 percent of registrants. Figure 6.2 and Figure 6.3 show that houses sell for a similar amount in the treatment and control groups for this subset of registrants. Home value amount is estimated based on a number of public record data, such as documents filed at the county recorder's office.

Next, we conduct the same analysis except defining those on the side of a block experiencing a polling place change as the treatment group and those of the side of a block keeping the same polling place as the control group. The one notable difference is that, because we have substantially fewer observations, Figure 6.4 sometimes reveals larger percent differences than were observed in Figure 6.1. Once again Figure 6.5 and Figure 6.6 show housing prices are comparable on the side of blocks that do and do not experience shocks for the 12 percent of registration records for which that information is available. Unlike with the relative distance, we also expect to observe similar voting patterns in the treatment and control groups in the previous election before the shock was realized, which Figure 6.7 shows is the case.

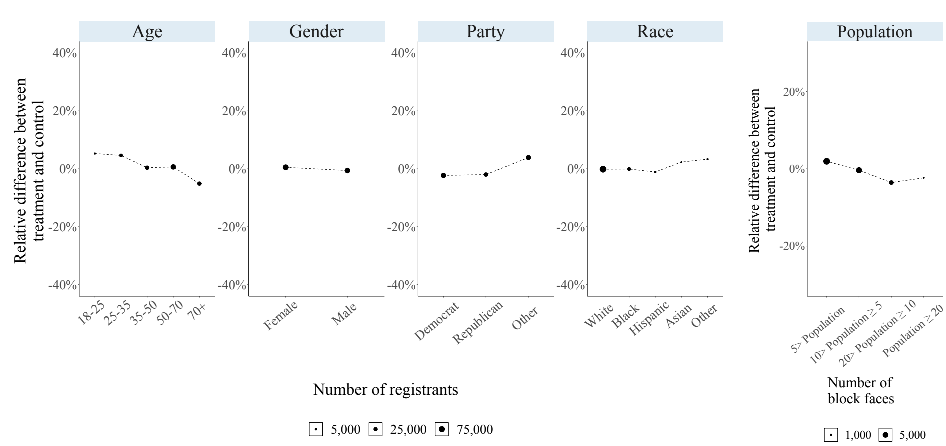


Figure 6.1. We see the age, gender, modeled partisanship, modeled race of registrants, as well as indicators representing that the total number of registrants are similar on either block face. (Distance)

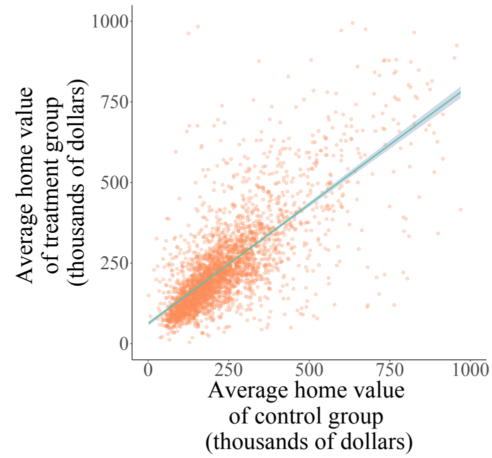


Figure 6.2. We see that the home prices of registrants on either block face are similar. (Distance).

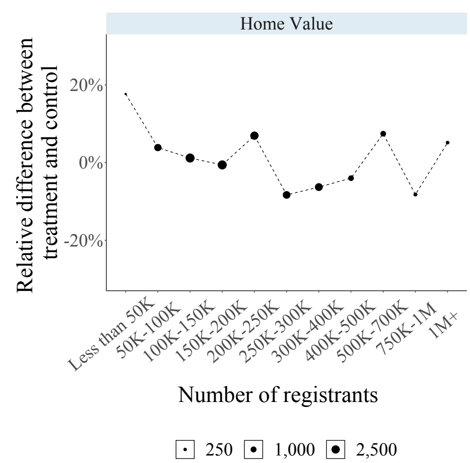


Figure 6.3. We see that the home prices of registrants on either block face are similar. (Distance)

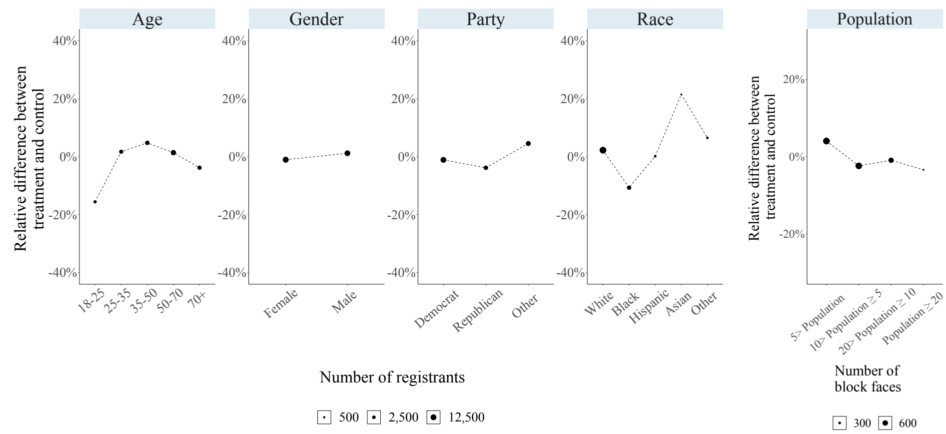


Figure 6.4. We see the age, gender, modeled partisanship, modeled race of registrants, as well as indicators representing that the total number of registrants, are similar on either block face, although the percent difference can be larger because the sample size is smaller. (Shocks)

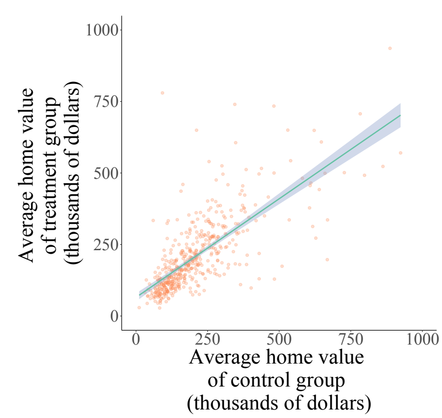


Figure 6.5. We see that the home prices of registrants on either block face are similar. (Shock)

Figure 6.6. We see that the home prices of registrants on either block face are similar. (Shock)

In Figure 6.8 we show the effect of relative distance for all blocks where the additional distance traveled by the treatment group ranges from 0 to more than 2 miles. When we include all voters we see that the average distance is quite small and the effect of relative distance is around 2 percentage points. While we see a more pronounced effected of relative distance for those voters who live on a block with a relatively large difference between the distances traveled by the close and far block faces, this might be attributable to other characteristics of these voters which differentiate them from the rest. For example, they tend to be more rural and more white.

Our baseline analysis only uses blocks in which all registrants on the block reside within 0.3 miles of one another. Figure 6.9 show our results when we instead apply a cutoff of 0.1 miles or 0.5 miles. Changing the cutoff affects our measures of uncertainty given that we are decreasing the sample size when we employ a cutoff of 0.1 and increasing the sample size when employ a cutoff of 0.5. And there is some sensitivity of the estimates within certain specific states. But we reach nearly identical conclusions based on the pooled analysis when we apply a cutoff of 0.1 or 0.5 instead of 0.3.

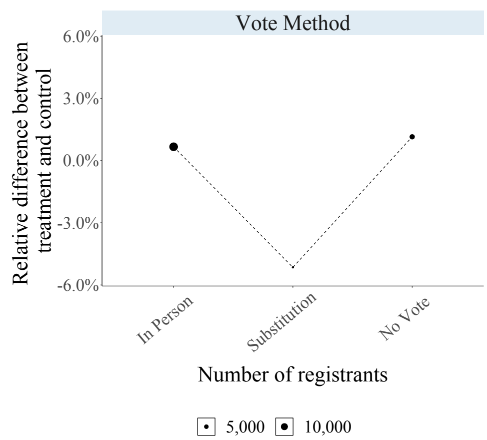


Figure 6.7. We see that the historical voting patterns in the 2012 presidential election of registrants on either block face are similar, with registrants experiencing a shock being, if anything, less likely to vote by substitution than registrants who did not. The roughly 5 percent lower usage of substitution voting reflects that 13.1 percent of registrants who experienced a shock voted using a substitution method in 2012 as compared to 13.8 percent of registrants who did not.

Our baseline analysis only includes registrants who have been registered to vote at the same address of registration since 2012. Figure 6.10 compares the effect of relative distance we estimate on this sample (left panel) to the the effect of relative distance we estimate when all 2016 registrants can be included in our sample (right panel). In this new analysis in-person voting decreased by 1.4 ± .3 p.p. (compared to 1.5 ± .4 in the previous analysis), and substitution voting increased by 1.1 ± .2 p.p (compared to 1.6 ± .6 in the previous analysis). Thus, we reach similar conclusions based on the pooled analysis whether or not new registrants are included in the analysis.

6.2 Types of shocks

Polling-place assignment changes can occur when precinct boundaries are redrawn or to meet the changing constraints placed on local election officials, e.g. when new regulations render a former site unfit. Broadly, polling places are 1) removed and consolidated within jurisdiction, 2) added to a jurisdiction and 3) moved to different locations where the total number within a jurisdiction is unchanged. To examine the types of shocks experienced by voters in this analysis, we describe each county in the dataset as either removing, adding, or not changing the total number of polling places from 2012 to 2016. Figure 6.1 shows both the percentage of counties and the percentage of voters living in those counties for each type of change to the number of polling places at the county

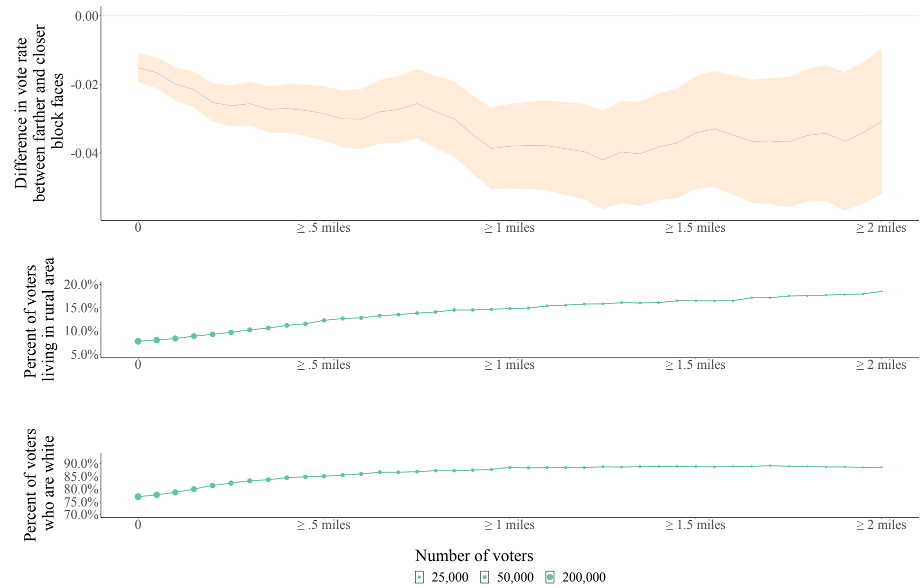


Figure 6.8. As the difference in distance between the closer and farther block faces grows, so does the effect of distance. However, the population also changes with distance.

level. We find that in our data the majority of voters who experience a shock live in a county where polling places were moved. To inspect county level changes, we consider a polling place to be a precinct name, assuming that within a county precinct names are not duplicated. This definition however may be likely to over-estimate the number of polling places as several precincts are often assigned to the same polling place.

	Percent of counties	Percent of registrants
Polling places removed in 2016	14.6	21.6
Polling places added in 2016	40.4	55.3
Same number of polling places in 2016 and 2012	44.9	23.2

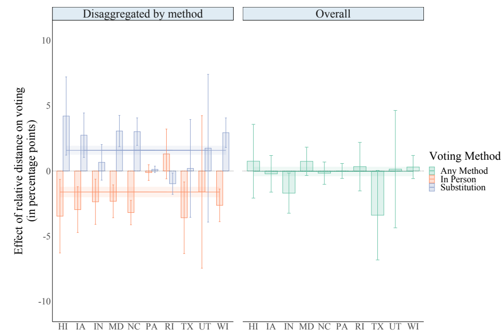
Table 6.1. Different types of shocks and the percentage of counts and voters affected.

Each of these events might influence voting differently. For example, consolidations leave fewer polling places for voters, increasing the distance for some voters to reach the polling place. They might also increase the cost of voting via longer lines and less staff members at the polling place. Alternatively, new polling places might reduce some costs, such as distance and time at the polling place. Analysing these distinct events would likely uncover different effects for each.

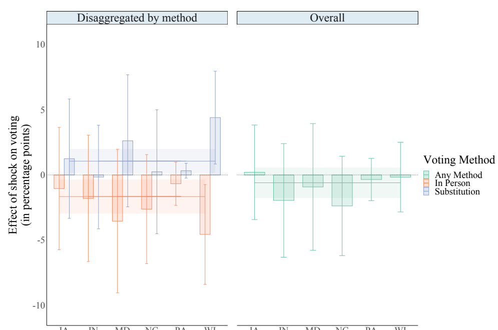
Here, we describe shock broadly, as having one's polling place assignment changed. While this description does not differentiate between the nuanced experiences of each type of change, it captures the common event of experiencing a search cost, where a voter must now determine where to vote. We see that this cost reduces in-person and increases substitution voting, respectively, even when the type of shock is not considered and when roughly 55% of all voters live in counties with polling-place additions.

When voters experience polling-place assignment changes they might receive a mailer describing the new assignment. In some cases this mailer might serve as an advertisement for substitution voting methods, describing how voters can pursue alternatives. The effect that we see might in part be the result of substitution methods being advertised to voters. However, we do see a reduction in

(a) Distance 0.1



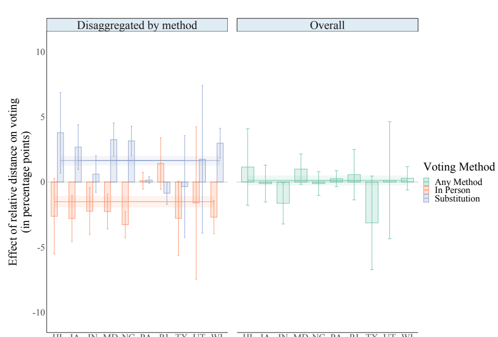
(b) Distance 0.5



(c) Shock 0.1

(d) Shock 0.5

Figure 6.9. We observe similar patterns when applying a smaller (0.1 miles) and larger (0.5 miles) cutoff than we apply in our baseline analysis



(a) Baseline Analysis

(b) All Registrants

Figure 6.10. We find similar patterns when estimating the effect of relative distance on all registrants as we do in our baseline analysis, when we exclude registrants who have not been registered at the same address since 2012.

in-person voting across states with different access to substitutes, different forms of substitutions and likely different mailers.

6.3 State-Supplied Polling-Place Location Files

Public records requests were made to state election officials in all fifty states. In response to our requests, most states (31) provided partial records; the remaining states denied our request (1), failed to complete the request (10), did not have applicable records because of the prevalence of voting by mail (2), or were unable to fulfill the request because relevant polling place location information was only available at the county level (6). Of the states which provided partial records 18 were complete enough to be potentially included in the analysis, but the number of states with a sufficient number of voters varied by experimental design. Each address in the state-supplied files is geocoded, according to a process described in Section 6.4.

For six states it is not possible to query a single state election official as polling-place locations are aggregated at the county level. Instead, for the largest of these states (Texas) we filed public records requests with county clerks in each of 254 counties. We received partial records from 142 counties, but were only able to recover reliable polling-place location data in 106 of these.

6.4 Filtering geocodes

State-supplied polling-place files describe each polling-place location with address descriptors such as street name, city, zip code and a textual place description. Our goal is to generate a *geocode* for each location, using Google's geocoding API [4]. We form two potential addresses to geocode for each location. The first address is the full street address which consists of a street number, street name, street descriptor (such as st. or ave.), as well as the city and state. The second address is the place address which consists of the place description, city and state. For example, a polling place might be described as the Local Elementary School and the resulting place address might be Local Elementary School, Doeville TX.

If the full street address is found to be precise and accurate according to the system below, its geocode is counted as valid and the associated latitude-longitude coordinates are used. If, however, the full street address fails either check, we instead attempt to geocode the place address. A geocode belonging to a place address must also be precise, but it only needs to pass the city and state accuracy tests.

Precision Each geocode is accompanied with metadata about the geocoding process. Here, we use a geocode's *location type*. There are four potential location types: rooftop, geometric center, range interpolated and approximate. If a geocode's location type is classified as rooftop we accept it immediately, while we immediately reject all geocodes classified as approximate.

If a geocode is classified with a location type of geometric center or range interpolated we accept it conditionally. We would like to accept all location types of sufficient granularity; however some subjectivity is required to discern between categories of geometric centers or interpolated ranges. We take a conservative approach and retain geometric centers or interpolated ranges which are tagged as an establishment, store, local government office or point of interest.

Accuracy Another item of metadata produced by the geocoding process is a formatted address of the geocode. We consider a geocode to be accurate if this formatted address matches the original address which was geocoded. To check accuracy we consider several components of an address, the street name, the street number, the city and the state.

4. https://developers.google.com/maps/documentation/geocoding/overview

Filtering Step	Number of potential voters remaining	
	2012	**2016**
Potential voters in valid voting jurisdictions (counties and polling places)	80,271,123	83,928,328
Filter potential voters with valid address	62,432,960	66,975,996
Filter potential voters with potential polling place assignment	41,756,035	41,756,035
Filter to registered and plausible voters	33,256,459	37,453,896
Filter to registrants who live on a block where all pairs of registrants live within .3 miles from one another	26,133,615	29,273,108
Filter to registrants who live on the same block in 2012 as in 2016	14,486,807	
Filter to registrants in **Distance analysis**:	252,428	
Filter to registrants in **Shock analysis**:	47,431	

Table 6.2. Here, we detail the data filtering steps followed to create final data set.

- **City and State** For each city and state we check that the formatted city and state match the original city and state. All geocodes whose formatted address city and state do not match the original city and state are considered inaccurate.
- **Street Address** To check that the street names and numbers match we first parse a street address into components using the Python software package USADDRESS.[5] This package standardizes address components so that, for example, 1st would be transformed to first. However, upon inspection, this standardization is not universally applied. Consequently, we performed a second hand-coded round of standardization where all numerical street names and descriptors were encoded in long-form (i.e. first and street or avenue), respectively. Finally, we check that a geocode's formatted address street number and name match the originals.
- **Evaluating Place Descriptions** If a street address is not considered to be precise and accurate, we instead geocode a location's place description. For example, a polling place might be described as Local Elementary School, Doeville TX. It is difficult to check the accuracy of a place description, as it will not contain an address, and descriptors returned from the geocoder might not match local descriptors. Instead of the full accuracy checks described above, for place descriptions we only check that the city and state of the place description's geocode match the original code. Thus, we only ensure that a geocode returned from a place description be precise, not that it be accurate.

6.5 Filtering to eligible registrants for analysis

Beginning with potential voters from the TargetSmart data file we conduct a series of filtering steps to produce the datasets used for the shock and distance analyses. These steps are described at a high level in Table 6.2 and in more detail in the following.

Initially, we begin with a set of **potential voters in valid voting jurisdictions (counties and polling places)**. Within the states included in the analysis there are some counties which hosted vote centers in either 2012 or 2016. In these counties registrants can cast a ballot at any vote center regardless of their registration precinct. This environment is in stark contrast to that which we assume in these analyses where voters must vote at their assigned precinct. Consequently, we remove all counties with vote centers in either 2012 or 2016 from the analysis.

Filter to potential voters with valid address As addresses are used to ensure the quality of a polling-place assignment, and in effect of relative distance experiment, we require that voters have valid addresses in order to computer their distance to the polling place. Several filtering steps are performed towards this end. We first drop all voters with missing address records. To compute a registrant's distance to their assigned polling place we rely on geo-coordinates and we filter out all registrants with missing geo-coordinates.

5. https://github.com/datamade/usaddress#readme

TargetSmart infers each voter's current address. We drop all voters whose current address does not match their registration address. Each geo-coordinate is accompanied with descriptors of the precision level at which it was recorded. For example, low precision levels are *Extrapolate* or *Zip Code* while a higher precision level is *Street*. A geo-coordinate at the level of *Extrapolate* refers to the closest known address to the original address and a geo-coordinate at the level of *Zip Code* refers to the centroid location of the zip code of the original address, while a geo-coordinate at the level of *Street* refers to the street address of the original address. We retain only those addresses which are geocoded at the level of *Street*.

Finally, each address is associated with a United States Postal Office (USPS) dwelling code. We retain only those residential addresses with a USPS code of High-rise, Building, or Apartment or Street Address. This removes all residences with a USPS code corresponding to a firm record, a general delivery area, a Post Office box or a rural route or highway.

Filter to potential registrants with potential polling place assignment To infer polling-place assignments for each voter we utilize both of our original data sources: the national voter file and the state-supplied polling-place files. Consider the three example voters in Table 2 and the two example polling places in Table 1. Voter-1 and Voter-2 would be assigned to PP-1 while Voter-3 would be assigned to PP-2. We filter out all voters who are assigned to a polling place which we could not geocode.

Once we have estimated a registrant's assigned polling place we calculate the distance between their home address and assigned polling place with the haversine formula which is provided with the Python package Geopy [6]. Because this matching process is vulnerable to errors, both in the original data sources and in the synchronization between them, we limit our analysis to registrants whose inferred assignments are no more than 25 miles from their home to remove egregious matching errors.

Filter to registered and plausible voters Next, we filter only to plausibly registered voters. That is we remove all unregistered voters and any voter who is marked as deceased. Hence, from this point on we refer to units in the analysis as registrants.

Filter to registrants who live on a block where all pairs of registrants live within .3 miles from one another Our analysis operates on blocks of registrants, where each block can be divided into two faces (see Figure 1 for an example). To identify eligible blocks of registrants for our analysis, we create a block identifier for each voter. This consists of all but the final two digits of the street number (e.g. 200 would be encoded as 2, 2100 would be encoded as 21), the street name, the street type, the city and state of a voter's residential address. For example, the two registrants, shown in Table 2 to be residing at 123 Main St. and 125 Main St. in Milwaukee, Wisconsin, respectively, share a block identifier.

To create a block identifier we require that an address consist only of numeric characters and that it be at least three digits long. The rational for the latter requirement is that in more rural locations where addresses are shorter there are examples of addresses with the same leading digit being on different blocks. For example, 80 Park Rd and 82 Park Rd might not be on the same block. We further ensure that all voters assigned to the same block live reasonably close by ensuring that no two voters with the same block identifier live more than 0.3 miles away from each other.

Filter to registrants who live on the same block in 2012 as in 2016 Finally, our analysis considers events that take place between the 2012 and 2016 elections. We therefore restrict our attention to registrants who resided on the same block in 2012 as in 2016.

Filter to registrants in analysis Finally, in both analyses we restrict our attention to blocks which meet the following eligibility criteria. In the TargetSmart data we observe instances where

6. https://geopy.readthedocs.io/en/stable/#module-geopy.distancehttps://geopy.readthedocs.io/en/stable/#module-geopy.distance

registrants with the same registration address or the same registration geocode are assigned to different precincts. We consider this to be an administrative error. One possible example of such an error is that a voter's registration is out-of-date and the record indicates a previous polling-place assignment. To prevent these errors from effecting our analysis, we filter out all blocks where multiple polling places are assigned to a single address or geocode. Additionally, we restrict our attention to those blocks where each block face has at least two voters.

Finally, for each analysis, we filter to all registrants still in the dataset who reside on a block which is present in both the filtered 2012 and filtered 2016 data. For the shock analysis this produces 47,456 voters. For the distance analysis this produces 249,309 voters.