Appendix

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

A.1 Additional Figures

Figure A.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 A.2 and Figure A.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 A.4 sometimes reveals larger percent differences than were observed in Figure A.1. Once again Figure A.5 and Figure A.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 A.7 shows is the case.

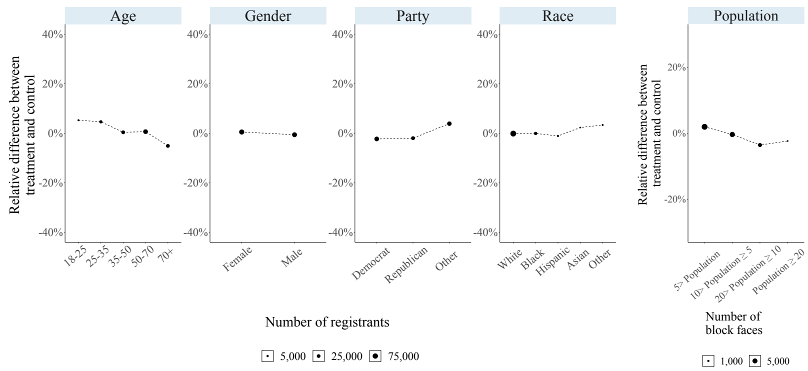


Figure A.1: We see the age, gender, modeled partisanship, modeled race of registrants, as well as indicators representing that the total number of registrants are similar on either block face. (Distance)

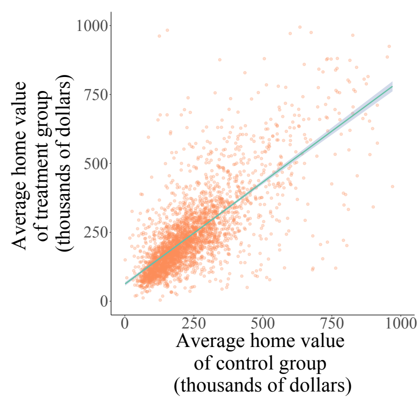


Figure A.2: We see that the home prices of registrants on either block face are similar. (Distance).

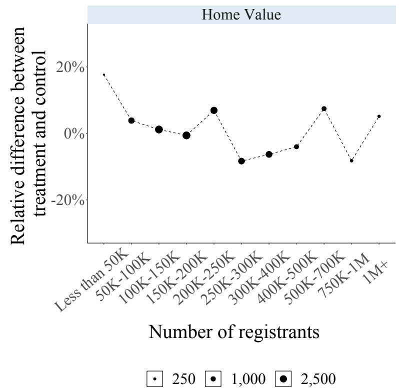


Figure A.3: We see that the home prices of registrants on either block face are similar. (Distance)

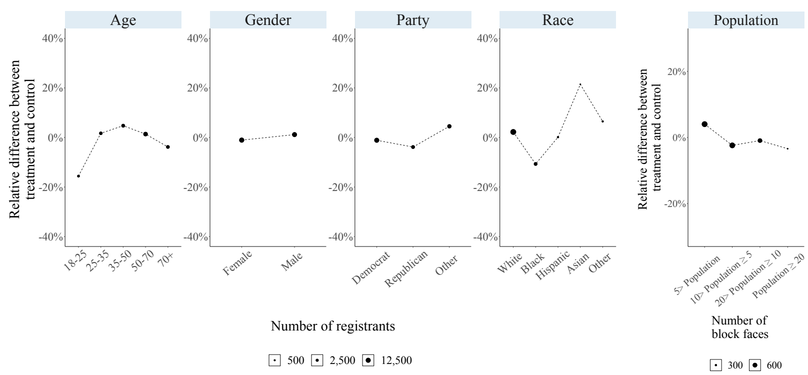


Figure A.4: We see the age, gender, modeled partisanship, modeled race of registrants, as well as indicators representing that the total number of registrants, are similar on either block face, although the percent difference can be larger because the sample size is smaller. (Shocks)

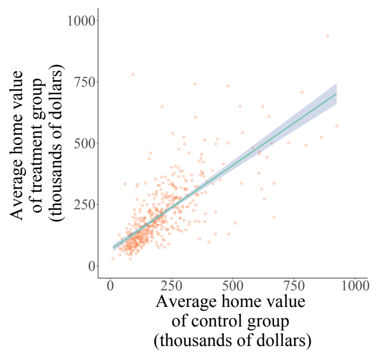


Figure A.5: We see that the home prices of registrants on either block face are similar. (Shock)

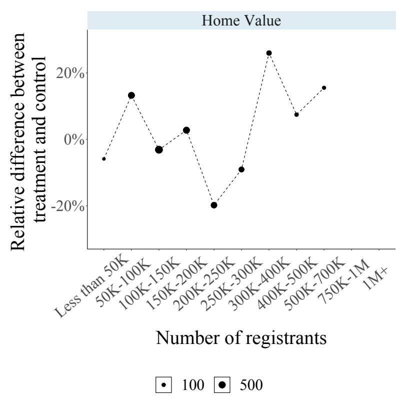


Figure A.6: We see that the home prices of registrants on either block face are similar. (Shock)

In Figure A.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 A.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.

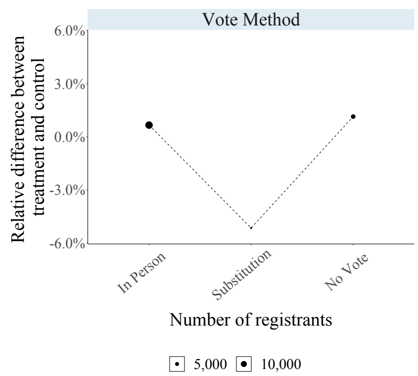


Figure A.7: We see that the historical voting patterns in the 2012 presidential election of registrants on either block face are similar, with registrants experiencing a shock being, if anything, less likely to vote by substitution than registrants who did not. The roughly 5 percent lower usage of substitution 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 A.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 $1.4 \pm .3$ p.p. (compared to $1.5 \pm .4$ in the previous analysis), and substitution voting increased by $1.1 \pm .2$ p.p (compared to $1.6 \pm .6$ in the previous analysis). Thus, we reach similar conclusions based on the pooled analysis whether or not new registrants are included in the analysis.

A.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 A.1 shows both the percentage of counties and the percentage of voters living in

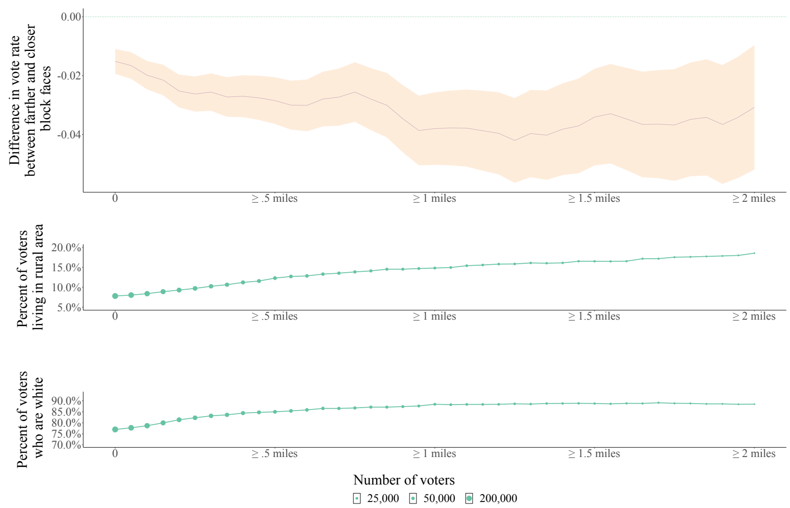


Figure A.8: As the difference in distance between the closer and farther block faces grows, so does the effect of distance. However, the population also changes with distance.

those counties for each type of change to the number of polling places at the county 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.

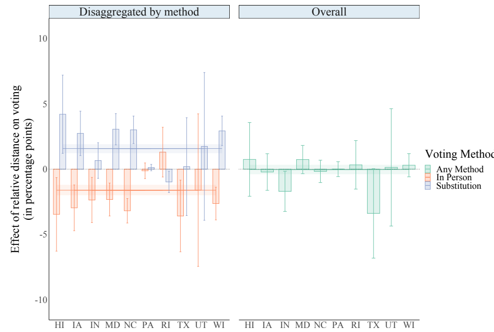
	Percent of counties	Percent of registrants
Polling places removed in 2016	14.6	21.6
Polling places added in 2016	40.4	55.3
Same number of polling places in 2016 and 2012	44.9	23.2

Table A.1: Different types of shocks and the percentage of counts and voters affected.

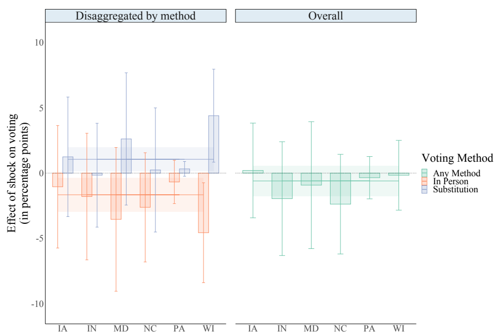
Each of these events might influence voting differently. For example, consolidations leave fewer polling places for voters, increasing the distance for some voters to reach the polling place. They might also increase the cost of voting via longer lines and less staff members at the polling place. Alternatively, 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

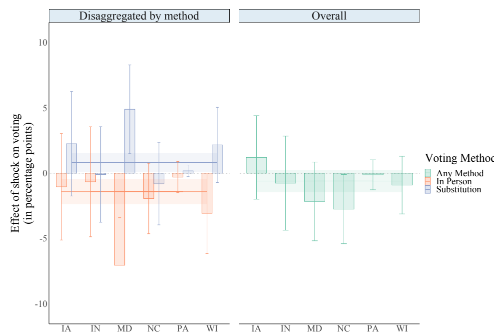
(a) Distance 0.1



(b) Distance 0.5

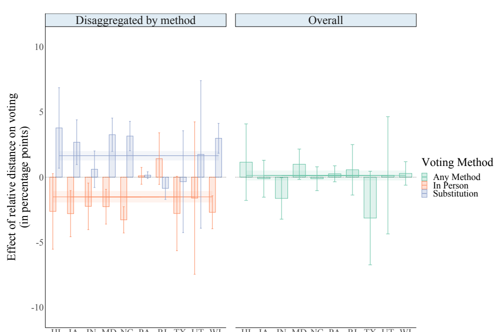


(c) Shock 0.1



(d) Shock 0.5

Figure A.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 A.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.

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

A.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 A.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.

A.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 [1]. 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.

[1] https://developers.google.com/maps/documentation/geocoding/overview

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.

- **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.[2] 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

[2]https://github.com/datamade/usaddress#readme

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 A.2: Here, we detail the data filtering steps followed to create final data set.

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.

A.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 A.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.

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

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 A.3: Example polling place file. Table 1 from the main text is copied here for the sake of exposition.

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 A.4: Example voter records. Table 2 from the main text is copied here for the sake of exposition.

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 A.4 and the two example polling places in Table A.3. 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 [3]. 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. in the main text for an example). To identify eligible blocks of registrants for

[3] `https://geopy.readthedocs.io/en/stable/#module-geopy.distance`https://geopy.readthedocs.io/en/stable/#module-geopy.distance

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 A.4 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 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 47,321 voters. For the distance analysis this produces 252,428 voters.