Voting Among Siblings

By MICHAEL D. BLOEM, JOHN B. HOLBEIN, SAM IMLAY, AND JONATHAN SMITH*

Using millions of young adult siblings in the U.S., we detail three findings that quantify whether siblings influence one another to vote in national elections. First, and descriptively, younger siblings are 10 percentage points (50 percent) more likely to vote in their first eligible election when their older sibling votes in a prior election. Second, roughly one-third of this is causal, as determined by age-of-voting-eligibility thresholds in a regression discontinuity design. Third, the causal impact of voting goes in the other direction as well - younger siblings increase the probability of their older siblings voting in their early 20s by 12 percent.

Keywords: Voting; Civic Engagement; Family; Siblings

I. Introduction

In the United States, only two-thirds of the voting-age population cast a ballot in 2020—an election with record turnout. Among 18-24 year olds, only 51.4 percent voted. There is no shortage of research and explanations as to why people do (or do not) turnout to vote, including the role of the family. Much of the family-focused literature examines the parental attributes that lead children to become voters (e.g. Levin, 1961; Clarke, 1973). For example, parents' level of education and marital status are correlated with children's voting behavior (Weisberg, 1987; Verba, Schlozman and Burns, 2005).

In this paper, we examine whether the decision of individuals to vote affects their siblings' decisions to vote. This departs from the earlier literature on family and voting in two distinct ways. First, we focus on the role of siblings on one another, not parents on children. Siblings have been described as the "ultimate peer" that give us a window into how close companions can influence one another, even in the political process. Second, we focus on individuals' civic actions on their siblings' voting behavior instead of a family or individual's attributes or circumstances. Focusing on one sibling's act of voting provides a direct and clear connection to the outcome of interest—voting.

Siblings share more in common than most, which creates a challenge to estimating whether siblings impact each other's probability of voting. For example, siblings often attend the same schools, have the same neighbors, and are exposed to similar civic and political acts and discussions. Each of the similarities may impact a person's probability of voting, such that any common voting patterns may stem from those underlying similarities and not from siblinghood. These similarities in voting patterns among siblings are borne out in our data.

* Bloem: College Board. Holbein: University of Virginia. Imlay: College Board. Smith: Georgia State University. This paper represents the views of the authors and not their organizations or that of the data providers.

Using data on 3.9 million sibling pairs from educational assessment data linked with administrative voting records, we show that younger siblings are 10 percentage points (50 percent) more likely to vote in their first eligible election when their older sibling votes in a prior election. We are the first to demonstrate this exceptionally strong correlation between siblings' voting decisions. It is also accomplished with the accuracy of administrative voting records, which can be an improvement over survey results biased from social desirability (Karp and Brockington, 2005; Holbrook and Krosnick, 2010) and for a much larger sample than ever before.

To move beyond correlations, we use a regression discontinuity design that exploits age-of-voting-eligibility cutoffs. In the U.S., individuals must be 18 years old on or before election day to vote. For example, someone born on November 4th, 1990 was exactly 18 on election day in 2008 and eligible to vote. Conversely, someone born one day later—November 5th, 1990—was one day shy of being eligible to vote in the 2008 election. On average, the people born on the 4th versus the 5th should be the same on all other dimensions, thereby setting up our natural experiment. We then compare the voting decisions of the younger (and older) siblings of the people born on the 4th versus the 5th and test whether they vote at different rates from one another.

We first quantify the above thought experiment that serves as our first stage—exogenous variation in voting using age-of-voting eligibility birth dates. About 20 percent of people, including older and younger siblings, vote just after turning 18 in an election year (30 percent in presidential elections). Voting around age 18 then translates into an increased likelihood of voting around age 20 and 22 compared to those who did not vote around age 18. We are not the first to ask the question of whether voting is habit-forming (e.g., Fujiwara, Meng and Vogl, 2016; Gerber, Green and Larimer, 2008). Most similar to our paper, Coppock and Green (2016) use discontinuities in voting by birth date relative to the age of eligibility. However, they only have access to administrative voting data, not a base sample that includes non-voters, so they rely on density jumps at the aggregate birth date level. We expand on their work with our education dataset that serves as a base dataset with non-voters, along with data from all states, not just the 17 they use.

Next, we test for spillovers from the older sibling to the younger sibling. An older sibling may serve as a role model or source of information that may induce the younger sibling to vote. But that requires the older sibling to vote. We find that younger siblings are 3.3 percentage points (14.4 percent) more likely to vote in their first eligible election if their older sibling voted around age 18. This causal effect is roughly one-third of the correlation previously mentioned. It is primarily driven by older siblings who voted around age 18 during a presidential election and for siblings of the same sex.

In the opposite direction, we also find that younger siblings who vote around age 18 increase the probability of the older sibling voting. Older siblings are 3.6 percentage points (12.3 percent) more likely to vote when their younger sibling votes around age 18 than older siblings whose younger sibling did not vote. This is one of the few papers to demonstrate spillovers from younger to older siblings in any field (e.g., Black et al., 2021; Karbownik and Özek, 2023; Humlum, Morthorst and Thingholm, 2024).

II. Data, Matching, and Samples

This study primarily uses two datasets: (1) sibling data derived from the universe of PSAT, SAT, and Advanced Placement (AP) takers in the high school graduation cohorts of 2004-2012; and (2) national voting records. We describe each in turn.

A. *Sibling Data*

The base data for this study are individual-level administrative records for all PSAT, SAT, and AP takers from the 2004-2012 high school graduating cohorts (21.3 million students). The SAT is one of two college entrance exams considered in admissions and program placement by thousands of colleges across the U.S. Approximately 1.3 million students per cohort take the exam. The PSAT is an exam taken prior to the SAT. It qualifies students for scholarships and college outreach and is the qualifying exam for the National Merit Scholarship. It has broader reach than the SAT and is usually taken during the sophomore and/or junior year of high school. Students take AP exams at the end of a school year, typically after taking a corresponding AP course. Performing well on the exams can earn students college credit while still in high school.

These testing data include a sibling identifier, which has been used in several other papers (e.g. Goodman et al., 2015; Gurantz, Hurwitz and Smith, 2020; Altmejd et al., 2021).[1] We exclude twins and multiples because our research question requires older and younger siblings. This leaves over 1.4 million sets of siblings, which corresponds to almost 3 million individuals. The vast majority of these siblings are sets of two.[2]

The testing data also contain several other pieces of information we rely on. First, it contains individual's exact date of birth, which we rely on in our regression discontinuity design. It also tells us the birth order within siblings and the disparity in ages. Second, it includes test scores for each exam. Third, it contains basic demographics, such as sex, race/ethnicity, self-reported parental income/education, and the high school attended.

B. *Voting Records*

In the United States, each state collects and reports its own voting data, but all states publicly detail whether (but not for whom) each registered voter votes. We obtained nationwide voting data from the Data Trust, LLC—one of the many vendors in this space. The Data Trust combines and standardizes data from each state's election governing body on the tens of millions of people who vote in biannual national elections from 2004 to 2016.

The data record each individual's name, date of birth (DOB), state of residence, and voter-turnout status in various primary and general elections. Thus, for someone who was 18 years of age by November 2004, we observe their voting history between 2004

[1] Generally speaking, siblings are defined as students with the same last name and home address. See Goodman et al. (2015) for more details and limitations.

[2] Need to explain why and perhaps how it compares to national statistics.

and 2016.[3] We collapse the voting records into a single observation per person—as opposed to a single observation per person-state—based on name and DOB. This process is detailed in Appendix B.

C. *Matching*

We match the education records to the voting records using students' name and date of birth (DOB). We limit the dataset to individuals who live in the 50 states or D.C., so we remove international test takers, students from U.S. territories, and the very small number of students missing a DOB.

We start by matching unique name and DOB combinations from each dataset and then employ a series of fuzzy matching methods, as described in Appendix B. To indicate potential false matches (and non-matches), we also construct a series of indicators—flags for common names, partially missing DOB in the voting data, and females (who are more likely to change their last name)—which we later use in robustness tests.

Importantly, an individual in the testing data that does not match to the voting data is considered a non-voter. Because of our conservative and imperfect matching process, our aggregate voting rates in each election are slightly below the true voting rates for people of a similar age. However, in a series of validation exercises in Appendix B, we show that the patterns of voting rates in our matched dataset follow those of nationally available statistics by age, race/ethnicity, and election.

D. *Analytic Samples*

The summary statistics for the educational data, matched to the voting records are in Table 1. The first column is the full sample of test-takers, including over 23 million siblings and non-siblings while the second column is the sample of 6.5 million siblings, which is the focus of the paper.

The top portion of the table shows students' demographics and PSAT scores. The sample is slightly more likely to be female and 57 percent White. The average PSAT score is 91.

The first eligible election is displayed in the second panel. There are fewer people who are first eligible in 2004 because only some of the 2004 high school cohort are eligible for that election and almost all of the 2005 cohort's first eligible election is 2006.

The last panel shows shows the means of our outcome variables. 16.8 percent of students voted in their first eligible election. This is roughly the average between the next two outcomes—voted in the first presidential election and the first midterm election. 43.3 percent of students ever vote through 2016, which equates to a mean of 0.924 votes in general elections and only 0.166 votes in primary elections.

The sibling sample follows similar patterns but a few noticeable differences. For example, the sibling sample is more likely to be White and have higher PSAT scores. Their voting rates are also a bit higher than the overall sample.

[3] Voters can be removed from a state's records. Removals vary over time and by state but do not impact most voter records. Our analyses focus on recent elections where voters have not had a chance to be removed.

TABLE 1—INDIVIDUAL LEVEL SAMPLE SUMMARY STATISTICS.

	Full sample (1)	Siblings sample (2)	Older sibling RD sample (3)	Younger sibling RD sample (4)
Individual characteristics				
Female	0.527	0.518	0.519	0.518
White	0.568	0.671	0.680	0.659
Black	0.139	0.079	0.077	0.079
Asian	0.068	0.073	0.070	0.079
Hispanic	0.157	0.125	0.122	0.132
American Indian	0.008	0.006	0.006	0.006
Other race	0.041	0.036	0.036	0.037
Missing race	0.027	0.016	0.015	0.014
PSAT score	91.411	97.671	98.000	97.902
Missing PSAT score	0.188	0.146	0.143	0.144
Parental educ: no college	0.113	0.102	0.102	0.103
Parental educ: college but no BA degree	0.129	0.136	0.140	0.138
Parental educ: college with BA degree	0.287	0.406	0.412	0.410
Parental educ: missing	0.471	0.356	0.346	0.350
Parental income: less than $50,000	0.129	0.111	0.118	0.107
Parental income: $50,000 - $100,000	0.130	0.168	0.182	0.159
Parental income: greater than $100,000	0.089	0.138	0.143	0.146
Parental income: missing	0.652	0.583	0.558	0.588
First eligible election				
2004	0.106	0.087	0.128	0.023
2006	0.182	0.182	0.212	0.155
2008	0.201	0.234	0.288	0.190
2010	0.211	0.237	0.207	0.276
2012	0.215	0.192	0.159	0.226
2014	0.084	0.068	0.007	0.130
Voting outcomes				
Voted in 1st election	0.168	0.196	0.214	0.193
Voted in 1st presidential election	0.259	0.296	0.296	0.316
Voted in 1st midterm election	0.082	0.098	0.101	0.099
Ever voted	0.433	0.501	0.517	0.501
Total general election votes	0.924	1.090	1.195	1.012
Total primary election votes	0.166	0.204	0.213	0.182
Number of individuals	23,362,841	6,490,550	225,556	181,994

Notes: The sibling sample includes all individuals with a sibling in the full sample. The older sibling RD sample includes individuals in sibling pairs that include the oldest child in the family with age differences between 9 and 48 months and where the older sibling turned 18 near the 2004, 2008, or 2012 elections. The younger sibling RD sample includes individuals in sibling pairs from families with 2 children with age differences between 9 and 48 months and where the younger sibling turned 18 near the 2004, 2008, and 2012 elections.

Many of our analyses focus on individuals who turn 18 close to an election. Table 1 also includes summary statistics for those who turn 18 within 30 days of a presidential election (i.e., 2004, 2008, 2012). The sample sizes drop precipitously but most of the

means are similar to the full sibling sample.

III. Descriptive Statistics and Analyses

We begin by documenting several new facts about the voting habits of siblings in Table 2. First, 67 percent of two-sibling households have neither sibling voting in the first election where both siblings are eligible to vote. This is not surprising given the low voting rates of young adults detailed in Table 1. Only one sibling votes 24.7 percent of the time and the older sibling is only slightly (one percentage point) more likely to be the voter than the younger sibling. The remaining 8.2 percent of the time both siblings vote. This implies that both siblings vote 25 percent of the time, conditional on at least one sibling voting.

TABLE 2—SIBLING VOTING RATES IN YOUNGEST SIBLING'S 1ST ELECTION.

	Sibling pair level	Family level	
	All sibling pairs (1)	Families with 2 children (2)	Families with 3 children (3)
Sibling vote rates			
Younger sibling voted	0.199	0.200	0.203
Middle sibling voted	–	–	0.207
Older sibling voted	0.211	0.211	0.231
Number of siblings voted			
Zero	0.670	0.671	0.575
One	0.249	0.247	0.266
Two	0.081	0.082	0.121
Three	–	–	0.038
Number of sibling pairs	3,886,316	–	–
Number of families	–	2,603,650	356,858

Notes: The sample in column 1 includes all siblings at the sibling pair level. The sample in columns 2 and 3 includes all siblings at the family level for families with 2 children (column 2) or 3 children (column 3).

Second, three-sibling households follow similar patterns with 57.5 percent of the trios not voting at all. Conversely, only 3.8 percent of the trios all vote.

Third, the correlation coefficient among pairs of siblings voting is about 0.25, which is relatively modest and implies that individuals within families do not typically have the same voting habits. However, there are some correlates that predict whether both siblings vote (Table A1), including relatively higher PSAT scores, parental income and education, and if both siblings are males. These are the same predictors of whether individuals vote in the data.

A. *Correlation Between Siblings' Voting*

We examine whether younger siblings are more likely to vote when their older sibling votes with the following descriptive regression:

$$v_{ie}^{y} = \alpha_0 + \alpha_1 v_{ie'}^{o} + X_i' A + E + \varepsilon_{ie} \tag{1}$$

The binary outcome v_{ie}^{y} is whether the individual i, who is a younger sibling, denoted by the superscript y, votes in election e. We are primarily interested in the coefficient α_1 on whether the older sibling, denoted by the superscript o, votes in election e'. For some analyses and specifications we consider the same elections for the older and younger siblings (i.e., $e = e'$). We include election year-by-state fixed effects E and, in some specifications, we control for the younger sibling's attributes (e.g., sex, test scores, high school), denoted X_i'.

The top panel of Table 3 shows that if older siblings vote in their first eligible election then their younger siblings are 11.3 percentage points more likely to vote in their first eligible election than if older siblings do not vote. Adding in controls for sex (column 2), PSAT scores (column 3), and high school fixed effects (column 4) slightly attenuates that relationship to 10 percentage points. The mean voting rate for younger siblings whose older sibling did not vote is roughly 20 percent, so 10 percentage points represents a 50 percent increase off of the base rate.

TABLE 3—DESCRIPTIVE RELATIONSHIP BETWEEN SIBLING VOTING PATTERNS.

	Outcome = younger sibling voted in 1st election			
	(1)	(2)	(3)	(4)
Older sibling voted in 1st election	0.114***	0.111***	0.108***	0.100***
	(0.001)	(0.001)	(0.001)	(0.001)
Outcome control mean	0.199	0.199	0.199	0.199
Number of sibling pairs	2,575,240	2,575,240	2,575,240	2,573,041
Older sibling voted in younger sibling's 1st election	0.175***	0.171***	0.168***	0.161***
	(0.001)	(0.001)	(0.001)	(0.001)
Outcome control mean	0.199	0.199	0.199	0.199
Number of sibling pairs	2,575,240	2,575,240	2,575,240	2,573,041
Election year by state fixed effects	✓	✓	✓	✓
Sibling demographics		✓	✓	✓
Sibling PSAT scores			✓	✓
Younger sibling high school fixed effects				✓

Notes: The samples includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months. Robust standard errors are reported.

The second panel of Table 3 mimics the top panel but the older siblings' voting variable is changed to whether the older siblings vote in the younger siblings' first eligible election. This election is often a few years after the older sibling's initial election, though sometimes it is the same. The last column shows that when older siblings vote in the younger siblings' first eligible election, the younger siblings are 16.1 percentage points (roughly 80 percent) more likely to vote than those whose older siblings did not vote.

The first two sub-panels of Table A2 display the same analyses but for the subsample of older siblings who are eligible to vote around presidential elections. Results are largely unchanged.

IV. Voting Begets Voting

This section shows that voting around the age of 18 leads to voting in future elections. This serves three purposes. First, we set up our regression discontinuity design that we use moving forward. The probability of voting, which discontinuously jumps when someone turns 18, serves as our first stage. Second, whether someone votes in downstream elections that take place well after age 18 helps us understand what it means to be treated (i.e., vote around age 18) and some potential mechanisms. Third, whether voting is habit forming is an interesting research question in itself (e.g. Coppock and Green, 2016; Fujiwara, Meng and Vogl, 2016; Gerber, Green and Larimer, 2008).

A. *Method - Regression Discontinuity*

We estimate the regression discontinuity below:

$$v_{ie} = \beta_0 + \beta_1 Eligible_{ie} + f(age_i) + X_i'B + E + \nu_{ie} \tag{2}$$

As before, the outcome v_{ie} is whether individual i votes in election e. age_i is the running variable, which is the exact date of birth, re-centered at zero relative to the closest election to i's 18th birthday. As an example, November 4th was election day in 2008. Someone born on November 3rd 1990 would have $age_i = -1$ and be eligible to vote in the 2008 election. Someone born on November 5th 1990 would have $age_i = 1$ and be ineligible to vote in the 2008 election. $f(.)$ is the functional form of the running variable, which is most often $f(age_i) = age_i$.

We are primarily interested in the estimate of β_1, which is the coefficient on $Eligible_{ie}$. This binary variable equals one if individual i is eligible to vote in the election closest to his or her 18th birthday and zero otherwise ($Eligible_{ie} = 1$ if $age_i \leq 0$). β_1 represents the discontinuous jump in the probability of voting relative to the trend in the running variable age_i. We always control for election fixed effects E and some specifications include individual-level controls X_i

We consider several outcomes, including voting in the election around the age of 18, followed by voting in subsequent elections. After the election around age 18, all the students become eligible to vote, despite using the same instrument $Eligible_{ie}$. Therefore, β_1 is a reduced-form coefficient for outcomes after election e.

Our primary specification uses a bandwidth of 30 days. That implies that we use the small subset of people born in October and November of an election year. The specification only uses the linear version of age, which is graphically justified, and uses robust standard errors. Results are not sensitive to any of these alternative specifications are still presented in the appendix.

Support for the Regression Discontinuity

Date of birth is unlikely to be manipulated but there are still two lingering concerns. First, date of birth determines whether someone takes the PSAT, SAT, or AP. This may very well be true for age-of-school entry, which is most frequently around August or September, but not around early November.

Second, people in the U.S. are less likely to be born on weekends. This creates lumpiness on certain birth dates, including right around the voting eligibility age/date (see Figure A1 and Figure A2). This can lead to some imbalance in some election years, but these weekend dips are visually obvious (2 out of 7 days) in election and non-election cohorts.

In Table A3, we estimate covariate balance tests that provide support for the research design. Specifically, we regress one covariate at a time (e.g., sex, PSAT scores, etc.) on the running variable and the eligibility threshold. We do not find evidence that the students' characteristics differ from one another on either side of the eligibility threshold.

B. Results

Using all siblings, Table 4 reveals two important results. First, in column 1, those who are barely eligible to vote around age 18 during an election year are 21.5 percentage points more likely to vote than those who are just barely not yet age 18 and ineligible to vote. This implies that 21.5 percent of those who are barely age 18 vote since those who are not yet 18 do not vote at all (visually confirmed in Figure A3. The estimate is 31 percentage points and 9.3 percentage points in presidential and midterm election years, respectively.

The first column estimates serve as the first stage in future analyses. Sometimes the samples change when we focus on an older or younger sibling or a particular election, but the magnitude of the coefficients are roughly the same, as seen across panels in Figure A3. Additionally, the relatively large coefficient on presidential elections provides relatively more statistical power and so many of our main analyses focus on presidential elections.

Second, Table 4 shows that voting begets voting, which is visually confirmed in Figure 1. The reduced-form estimate on being eligible to vote from Equation 4 is 1.3 percentage points. This means that students who are just barely 18 before an election are 1.3 percentage points more likely to vote two years later in the next election than those who were just barely not 18. This is a 10 percent increase over the mean for voting rate for those who were still 17 at the time of the initial election.

To put that in context, a 10 percent increase in the probability of voting is not small compared to other voting interventions (CITES). More importantly, if we scale the 1.3

TABLE 4—EFFECT OF ELIGIBILITY AT AGE 18 ON OWN FUTURE VOTING, ALL SIBLINGS SAMPLE.

	First stage	Own future voting outcomes				
	Voted at age 18 (1)	Voted in election 2 years later (2)	Voted in election 4 years later (3)	Ever voted in the future (4)	General election votes (5)	Primary election votes (6)
Panel A: 18th birthdays near any election						
Eligible to vote at age 18	0.215***	0.013***	0.009***	0.013***	0.046***	0.016***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.007)	(0.003)
Outcome control mean	0.007	0.174	0.223	0.480	0.949	0.181
Observations	511,470	511,470	511,470	511,470	511,470	511,470
Panel B: 18th birthdays near presidential elections						
Eligible to vote at age 18	0.310***	0.021***	0.016***	0.014***	0.057***	0.022***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.009)	(0.004)
Outcome control mean	0.007	0.085	0.310	0.463	0.893	0.176
Observations	286,614	286,614	286,614	286,614	286,614	286,614
Panel C: 18th birthdays near midterm elections						
Eligible to vote at age 18	0.093***	0.004	0.002	0.012***	0.032***	0.009*
	(0.002)	(0.004)	(0.003)	(0.004)	(0.011)	(0.005)
Outcome control mean	0.006	0.289	0.112	0.502	1.021	0.189
Observations	224,856	224,856	224,856	224,856	224,856	224,856

Notes: The sample includes all siblings with an 18th birthday within 30 days of an election. Regressions include fixed effects for the election year that is closest to their 18th birthday. Bandwidth is 30 days. Robust standard errors are reported.

percentage points by the first column, which serves as a first-stage, our estimate suggest that voting around age 18—as opposed to being ineligible to vote—increases the probability of voting again two years later by 6 percentage points (35 percent).

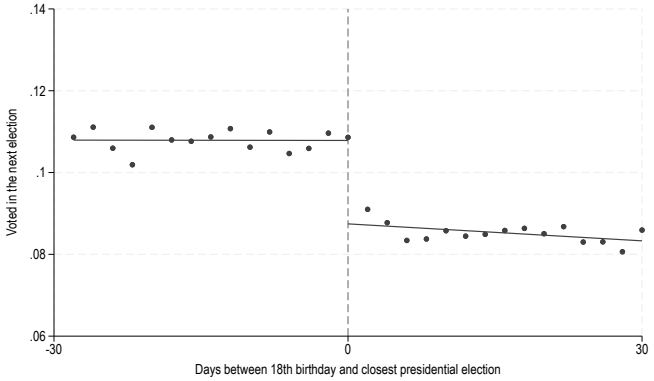
Being eligible to vote around 18 also increases the probability of voting in four years by 0.9 percentage points, the probability of ever voting by 1.3 percentage points, the number of general election votes cast by 0.046, and the number of primary election votes cast by 0.016 (see Figure A4). Overall, these results show that voting around 18 leads to future voting.

The previous results are driven mostly by students who turn 18-years-old around a presidential election. This can be seen in the next two panels of Table 4, which are split by students around age 18 during a presidential election and midterm election, respectively.

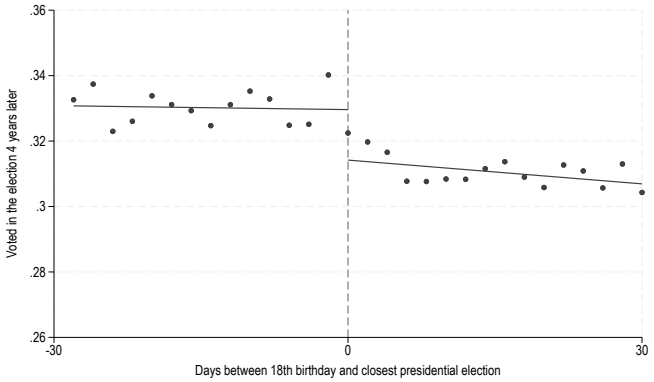
These results are insensitive to choice of bandwidth, weighting, and polynomial order of the running variable. These robustness tests are shown in Table A4. Table A5 (complemented by Figure A5) and Table A6 (complemented by Figure A6) show qualitatively similar results to Table 4 when using the full sample that includes non-siblings and the subsample of only older siblings, respectively.

FIGURE 1. THE EFFECT OF AGE 18 ELIGIBILITY ON FUTURE VOTING - ALL SIBLINGS SAMPLE

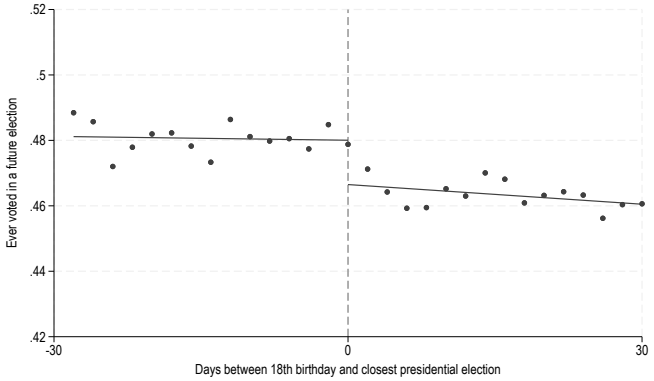
(a) Voted in election 2 years later



(b) Voted in election 4 years later



(c) Ever voted in future election



Note: This figure shows binned scatterplots of average voting outcomes by days between each sibling's 18th birthday and the presidential elections of 2004, 2008, and 2012. Each bin consists of 2 days.

V. Siblings' Spillovers - Methods

We use a regression discontinuity design to assess whether one sibling's voting impacts another sibling's probability of voting. For exposition's sake, the equations below are spillovers from the older sibling to the younger sibling but spillovers in the reverse direction use the same methods.

Building on Equation 2, we begin with a first-stage equation that estimates the discontinuous jump in the probability of voting when the older sibling turns 18 around an election with the following:

$$v^o_{ie} = \gamma_0 + \gamma_1 Eligible^o_{ie} + f(age^o_i) + X^o_i \Gamma + E + \upsilon_{ie} \tag{3}$$

The important distinction with Equation 2 is the superscript o on most of the variables that designates older siblings. The alternative is the superscript y for younger siblings, which appears in later equations.

We are interested in the coefficient γ_1, which measures the increase in the probability of the older sibling voting in election e when he or she turns 18 and becomes eligible to vote. Theoretically, nobody under age 18 should vote so the coefficient can also be interpreted as the voting rate for those who just turned 18.

Next, we consider the impact of the older sibling voting on the younger sibling's probability of voting through the following reduced-form equation:

$$v^y_{ie} = \delta_0 + \delta_1 Eligible^o_{ie'} + f(age^o_i) + X^y_i \Omega + E + \omega_{ie} \tag{4}$$

The outcome variable is whether the younger sibling voted in election e. This is the main distinction with the first-stage where the outcome was for the older sibling. But the right-hand-side running and eligibility variables remain about the older sibling. Additionally, the elections that the older and younger siblings need not be the same (i.e., $e \neq e'$).

Lastly, we combine the first-stage and reduced-form equations into a two-stage least squares analysis with the following second stage:

$$v^y_{ie} = \lambda_0 + \lambda_1 \hat{v}^o_{ie'} + f(age^o_i) + X^y_i \Lambda + E + \psi_{ie} \tag{5}$$

$\hat{v}^o_{ie'}$ is the instrumented endogenous variable from Equation 3. As such, λ_1 represents the causal impact of whether the older sibling votes in election e' on whether the younger sibling votes in election e.

A. *Identification and Implementation*

Identification comes from random variation in the exact date of birth and the election date. We have no reason to believe that someone born on November 5th versus November

6th are different from one another in any discernible way. We already demonstrated the lack of differences in the covariates Table A3. Columns 2 and 3 of this table repeats this exercise separately for older siblings and younger siblings, respectively, who turn 18 within 30 days of a presidential election and finds no concerns about manipulation of the running variable.

In practice, we use a bandwidth of 30 days on either side of the eligibility threshold and a linear functional form for the running variable. We later test the sensitivity of these decisions. We also use robust standard errors.

Lastly, we focus on the subsample of older (and younger) siblings who turn 18 around a presidential election (i.e., around November of 2004, 2008, 2012). Voter turnout among those just barely 18 years old in presidential elections is roughly 30 percent, as seen in Figure A3, compared to 10 percent in midterm election years. This larger first-stage helps with statistical power despite dropping siblings who turn 18 around the 2006 and 2010 elections. We show results using the full sample and the subsample of midterm elections in the appendix.

VI. Older Siblings' Spillovers

A. *Main Results*

Older siblings that vote around age 18 increase the probability that their younger siblings vote. Table 5 shows the reduced-form and second stage results, which are supported by Figure 2.

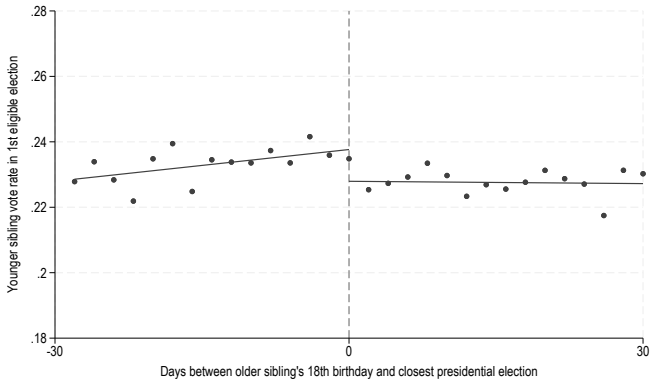
TABLE 5—OLDER SIBLING RD, 18TH BIRTHDAYS NEAR PRESIDENTIAL ELECTIONS.

	1st stage	Younger sibling voting outcomes						
	Older sibling voted at age 18 (1)	Voted in 1st election (2)	Voted in 1st pres. election (3)	Voted in 1st midterm election (4)	Ever voted (5)	Vote count (6)	General election vote count (7)	Primary election vote count (8)
Older sibling eligible to vote at age 18 (RF)	0.295*** (0.004) [5,665]	0.010** (0.005)	0.011** (0.005)	-0.001 (0.004)	0.004 (0.006)	0.042** (0.020)	0.032** (0.016)	0.010 (0.007)
Older sibling voted at age 18 (IV)		0.033** (0.016)	0.038** (0.018)	-0.004 (0.012)	0.014 (0.020)	0.142** (0.068)	0.108** (0.054)	0.034 (0.023)
Outcome control mean	0.009	0.229	0.293	0.099	0.501	1.227	1.039	0.188
Observations	112,778	112,778	112,778	112,772	112,778	112,778	112,778	112,778

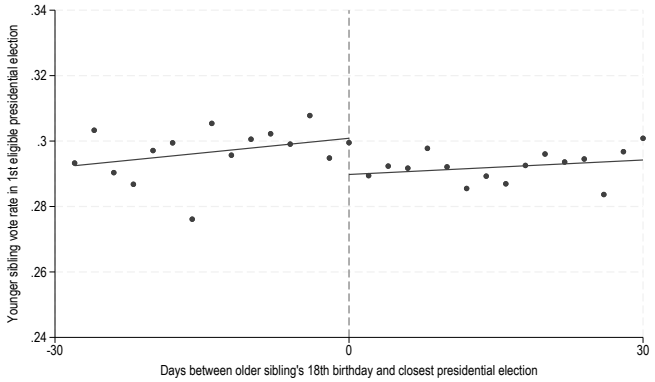
Notes: The sample includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months and where the older sibling turned 18 near the 2004, 2008, or 2012 elections. Regressions include fixed effects for the election year the younger sibling is first eligible for. Bandwidth is 30 days. First stage F-stat reported in brackets in column 1. Robust standard errors are reported.

FIGURE 2. EFFECT OF OLDER SIBLING'S AGE 18 ELIGIBILITY ON YOUNGER SIBLING'S VOTING

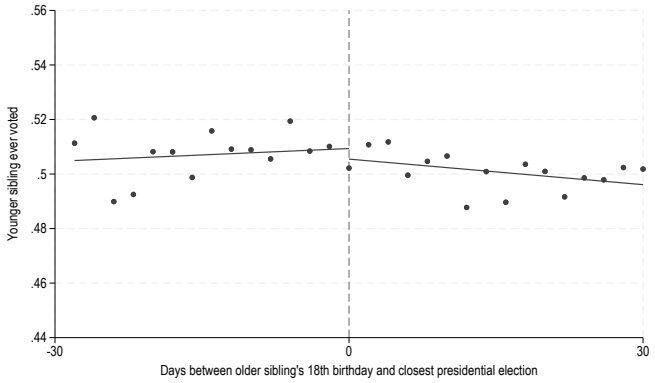
(a) Younger sibling voted in 1st election



(b) Younger sibling voted in 1st presidential election



(c) Younger sibling ever voted



Note: This figure shows binned scatterplots of average younger sibling voting outcomes by days between the older sibling's 18th birthday and the presidential elections of 2004, 2008, and 2012. Voting outcomes are residualized for fixed effects of the younger sibling's first election to match our estimating equation. Each bin consists of 2 days.

We find that younger siblings are 1 percentage point more likely to vote in their first election if their older sibling is just barely eligible to vote at age 18 compared to those with older siblings just barely ineligible to vote at age 18. This reduced-form estimate is small in magnitude but precisely measured. This translates to a 3.3 percentage point increase in the probability of voting when scaled by the first-stage (column 1) in the instrumental variables framework. To put that into perspective, 3.3 percentage points is a 14.4 percent increase relative to the 22.9 percent voting rate of the control mean. Putting it all together, younger siblings are 14.4 percent more likely to vote in their first election when their older sibling votes around age 18. This is visually confirmed in Panel A of Figure 2.[4]

The next several columns show different outcomes for the younger sibling. Columns 3 and 4 shows that the previous estimate is driven by younger siblings whose first election was a presidential election, which is four years after the older sibling turns 18. The next four columns show that these spillover effects seem to dissipate over time, since we estimate no statistically significant effects on whether the younger sibling ever voted in our data (column 5). However, the spillover effects still amount to a 0.14 increase in total votes cast by younger siblings (column 6), which is mostly driven by an increase in general election votes (column 7) as opposed to primary election votes (column 8).

The statistical significance of the main results is somewhat sensitive to choice of bandwidth, kernel weight, and polynomial order, as shown in Table A7. But the coefficients are consistently positive and of a similar magnitude when the outcome is either voted in first eligible election or first eligible presidential election. We also show no impact in the reduced-form estimates of the older sibling around age 18 in odd cohorts when there was no election (Table A8 and Figure A8).

Finally, we find that the effects are either muted or null when we include older siblings around age 18 in a midterm election (Table A9). This is because of the weaker first stage for midterm elections and perhaps because of the smaller effects on one's own future voting when they are 18 around a midterm election.

B. Heterogeneity and Potential Mechanisms

Table 6 explores potential mechanisms through several types of heterogeneous effects.

First, we explore whether siblings who are similar to one another on several attributes have larger effects. We find that same sex siblings are driving the results and there are no spillovers among opposite sex siblings (we show this visually in Figure A9).[5] Same sex siblings are 7.4 percentage points (33 percent) more likely to vote in their initial election when their older sibling previously voted around age 18. The difference in the reduced-form estimates between same sex and opposite sex siblings is statistically significant. We do not find strong evidence that age gaps between siblings are driving the results.

Second, we find that siblings who live in different states as one another have larger point estimates than those who live in the same state. The only way we know whether

[4] Additional outcomes that mirror the columns of Table 5 are shown in Figure A7.

[5] Table A10 splits by all four combinations of siblings sex combinations, which shows that sisters are the largest contributors to the spillover effects.

TABLE 6—OLDER SIBLING RD, HETEROGENEITY RESULTS.

	Outcome = Younger sibling voted in 1st election							
	Same sex sibling pairs (1)	Opposite sex sibling pairs (2)	Siblings in same state (3)	Siblings in diff. states (4)	Younger sibling age 18 at 1st election (5)	Younger sibling age 19 at 1st election (6)	Sibling age diff. 2 years or less (7)	Sibling age diff. between 2-4 years (8)
Older sibling eligible to vote at age 18 (RF)	0.022***	-0.003	0.005	0.020*	0.006	0.016*	0.008	0.011
	(0.007)	(0.007)	(0.006)	(0.010)	(0.006)	(0.008)	(0.006)	(0.007)
Older sibling voted at age 18 (IV)	0.074***	-0.009	0.017	0.061*	0.019	0.053*	0.026	0.037
	(0.023)	(0.024)	(0.020)	(0.031)	(0.020)	(0.027)	(0.021)	(0.023)
Outcome control mean	0.223	0.235	0.228	0.237	0.201	0.269	0.107	0.299
Observations	57,652	55,126	82,393	26,869	66,925	45,611	41,339	71,438

Notes: The sample includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months and where the older sibling turned 18 near the 2004, 2008, or 2012 elections. Regressions include fixed effects for the election year the younger sibling is first eligible for. Bandwidth is 30 days. Robust standard errors are reported.

the siblings are in different states is by the college they attend,[6] so we cannot isolate the role of geography.

VII. Younger Siblings' Spillover

A. *Main Results*

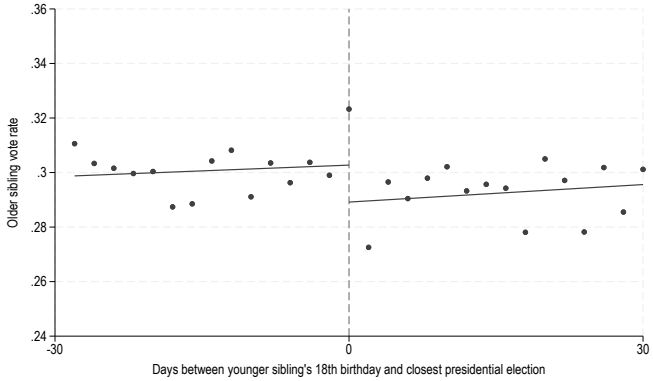
Younger siblings that vote around age 18 increase the probability that their older siblings vote. Table 7 shows these results, which are supported by Figure 3. We find that older siblings whose younger siblings are eligible to vote around age 18 are 1.2 percentage points more likely to vote than if their younger siblings are ineligible to vote around age 18. When this reduced-form effect is scaled by the first-stage (reported in column 1), we estimate that a younger sibling voting increases the probability that older siblings vote in the same election by 3.6 percentage points (12.3 percent). The results in column 2 of Table 7 are insensitive to choice of bandwidth, kernel weight, and polynomial order, as shown in Table A11.

Column 3 shows that this spillover from younger to older siblings do not persist to future elections. We estimate no statistically significant effect on older siblings vote rates in the election 2 years later. Unlike the spillover results in the other direction, we find that including sibling pairs with younger siblings turning 18 around midterm elections does not meaningfully change the results. This can be seen in the bottom panel of Table 7.

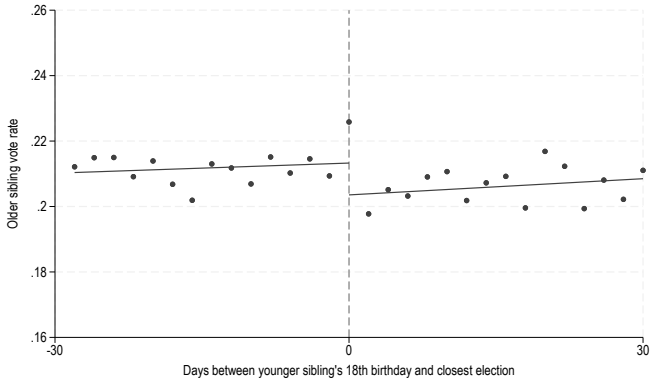
[6]Test taking data is linked to National Student Clearinghouse data, which is a near census of college enrollment in the U.S.

FIGURE 3. EFFECT OF YOUNGER SIBLING'S AGE 18 ELIGIBILITY ON OLDER SIBLING'S VOTING

(a) 18th birthdays near presidential elections



(b) 18th birthdays near any election



Note: This figure shows binned scatterplots of average older sibling vote rates in their younger sibling's 1st election by days between the younger sibling's 18th birthday and the election. Panel (a) includes sibling pairs where the younger sibling turned 18 near the presidential elections of 2004, 2008, and 2012. Panel (b) includes midterm elections. Voting outcomes are residualized for fixed effects of the election closest to the younger sibling's 18th birthday to match our estimating equation. Each bin consists of 2 days. Panel A includes same sex sibling pairs (brother and sister pairs). Panel B includes opposite sex sibling pairs.

B. Heterogeneity and Potential Mechanisms

Table 8 explores potential mechanisms through several types of heterogeneous effects. The first two columns show that the point estimate for opposite siblings is larger than for same sex siblings.[7] This is the opposite result from the spillovers from the older to younger sibling, although the difference in the reduced-form estimates between opposite and same sex siblings is not statistically significant at conventional levels.[8] Once again, we do not find strong evidence that age gaps between siblings is driving the results.

[7] We show this visually in Figure A10.

[8] Table A12 splits by all four combinations of siblings sexes and older sisters and younger brothers are the largest contributors to the spillover effects.

TABLE 7—YOUNGER SIBLING RD.

	First stage	Older sibling voting outcomes	
	Younger sibling voted in election closest to 18th birthday (1)	Voted in election closest to younger sibling's 18th birthday (2)	Voted in election 2 years later (3)
Panel A: Younger sibling's 18th birthday near presidential election			
Younger sibling eligible to vote at age 18 (RF)	0.332***	0.013**	-0.001
	(0.004)	(0.006)	(0.004)
	[5,567]		
Younger sibling voted at age 18 (IV)		0.040**	-0.003
		(0.018)	(0.013)
Outcome control mean	0.006	0.292	0.124
Observations	90,997	90,997	90,997
Panel B: Younger sibling's 18th birthday near any election			
Younger sibling eligible to vote at age 18 (RF)	0.223***	0.009**	-0.000
	(0.003)	(0.004)	(0.004)
	[1,457]		
Younger sibling voted at age 18 (IV)		0.042**	-0.001
		(0.017)	(0.017)
Outcome control mean	0.006	0.206	0.202
Observations	165,751	165,751	165,751

Notes: The sample includes sibling pairs from families with 2 children with age differences between 9 and 48 months and where the younger sibling turned 18 near the 2004, 2006, 2008, 2010, 2012, or 2014 elections. Regressions include fixed effects for the election year that is closest to the younger sibling's 18th birthday. Bandwidth is 30 days. Robust standard errors are reported.

VIII. Conclusion

This paper documents the impact siblings have on one another's probability of voting. We find evidence that older siblings influence younger siblings to vote and vice versa. As far as we know, we are the first to document sibling spillovers in voting.

Our findings are important beyond the somewhat narrow context of siblings and voting. We have no reason to believe that voting is the only action that siblings impact. However, most settings and actions are difficult to quantify any sibling spillovers. Our context benefited from rich data and an age-of-voting-eligibility discontinuity. Quantifying the impacts of policies and interventions are difficult enough and most evaluations likely cannot account for spillovers. Not accounting for spillovers has the potential to underestimate the policy or intervention's benefits or costs.

Additionally, voting is an important act of civic engagement in the U.S. and young adults cast their ballots at alarmingly low rates. We show that siblings can play a role in increasing voting rates. This knowledge can be exploited in policy and intervention, perhaps targeting particular family structures or specific individuals within families.

TABLE 8—YOUNGER SIBLING RD, HETEROGENEITY RESULTS.

	Outcome = Older sibling voted in election closest to younger sibling's 18th birthday					
	Same sex sibling pairs (1)	Opposite sex sibling pairs (2)	Older sibling age 18 at 1st election (3)	Older sibling age 19 at 1st election (4)	Sibling age diff. 2 years or less (5)	Sibling age diff. between 2-4 years (6)
Younger sibling eligible to vote at age 18 (RF)	0.008	0.019**	0.019**	0.010	0.011	0.015*
	(0.008)	(0.009)	(0.010)	(0.008)	(0.010)	(0.008)
Younger sibling voted at age 18 (IV)	0.025	0.056**	0.056**	0.031	0.036	0.042*
	(0.026)	(0.026)	(0.028)	(0.024)	(0.031)	(0.022)
Outcome control mean	0.290	0.295	0.295	0.291	0.291	0.293
Observations	45,961	45,036	36,545	53,917	35,696	55,301

Notes: The sample includes sibling pairs from families with 2 children with age differences between 9 and 48 months and where the younger sibling turned 18 near the 2004, 2008, or 2012 elections. Regressions include fixed effects for the election year that is closest to the younger sibling's 18th birthday. Bandwidth is 30 days. Robust standard errors are reported.

Finally, we think there is still much to learn about the role of the family in voting. We only focused on siblings but parents and more distant family members are worthy of further exploration. We also do not know how siblings impact one another once they're past their 20s, when voting rates are higher. Similarly, we were unable to explore many mechanisms in this paper and under what conditions our results are the strongest. We leave that to future research.

REFERENCES

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith. 2021. "O brother, where start thou? Sibling spillovers on college and major choice in four countries." *The Quarterly Journal of Economics*, 136(3): 1831–1886.

Bell, D'Wayne, Jing Feng, John Holbein, and Jonathan Smith. 2022. "Do STEM Students Vote?"

Black, Sandra E., Sanni Breining, David N Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth, and Marianne Simonsen. 2021. "Sibling Spillovers." *The Economic Journal*, 131(633): 101–128.

Clarke, James W. 1973. "Family structure and political socialization among urban black children." *American Journal of Political Science*, 302–315.

Coppock, Alexander, and Donald P Green. 2016. "Is voting habit forming? New evidence from experiments and regression discontinuities." *American Journal of Political Science*, 60(4): 1044–1062.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit formation in voting: Evidence from rainy elections." *American Economic Journal: Applied Economics*, 8(4): 160–188.

Gerber, Alan S, Donald P Green, and Christopher W Larimer. 2008. "Social pressure and voter turnout: Evidence from a large-scale field experiment." *American political Science review*, 102(1): 33–48.

Goodman, Joshua, Michael Hurwitz, Jonathan Smith, and Julia Fox. 2015. "The relationship between siblings' college choices: Evidence from one million SAT-taking families." *Economics of Education Review*, 48: 75–85.

Gurantz, Oded, Michael Hurwitz, and Jonathan Smith. 2020. "Sibling effects on high school exam taking and performance." *Journal of Economic Behavior & Organization*, 178: 534–549.

Holbrook, Allyson L, and Jon A Krosnick. 2010. "Social desirability bias in voter turnout reports: Tests using the item count technique." *Public opinion quarterly*, 74(1): 37–67.

Humlum, Maria Knoth, Marius Opstrup Morthorst, and Peter Rønø Thingholm. 2024. "Sibling Spillovers and the Choice to Get Vaccinated: Evidence from a Regression Discontinuity Design." *Journal of Health Economics*, 94(102843).

Karbownik, Krzysztof, and Umut Özek. 2023. "Setting a Good Example? Examining Sibling Spillovers in Educational Achievement Using a Regression Discontinuity Design." *Journal of Human Resources*, 58(5): 1567–1607.

Karp, Jeffrey A, and David Brockington. 2005. "Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries." *The Journal of Politics*, 67(3): 825–840.

Levin, Martin L. 1961. "Social climates and political socialization." *The Public Opinion Quarterly*, 25(4): 596–606.

Verba, Sidney, Kay Lehman Schlozman, and Nancy Burns. 2005. "Family Ties." *Social Logic Of Politics: Personal Networs As Contexts*, 95.

Weisberg, Herbert F. 1987. "The demographics of a new voting gap marital differences in American voting." *Public Opinion Quarterly*, 51(3): 335–343.

APPENDIX A - TABLES AND FIGURES

TABLE A1—DESCRIPTIVE RELATIONSHIP BETWEEN BOTH SIBLINGS VOTING AND FAMILY CHARACTERISTICS.

	Outcome = Both siblings voted in younger sibling's 1st election				
	(1)	(2)	(3)	(4)	(5)
Both siblings female (reference)					
Both siblings male	0.010***			0.009***	0.009***
	(0.000)			(0.000)	(0.000)
Older sibling female, younger sibling male	-0.004***			-0.005***	-0.004***
	(0.000)			(0.000)	(0.000)
Older sibling male, younger sibling female	0.005***			0.004***	0.003***
	(0.000)			(0.000)	(0.000)
Sibling's average PSAT score		0.001***		0.000***	0.000***
		(0.000)		(0.000)	(0.000)
Parental educ: no college (reference)					
Parental educ: college but no BA degree			0.012***	0.011***	0.010***
			(0.001)	(0.001)	(0.001)
Parental educ: college with BA degree			0.024***	0.020***	0.018***
			(0.001)	(0.001)	(0.001)
Parental income: less than $50,000 (reference)					
Parental income: $50,000 - $100,000			0.009***	0.008***	0.008***
			(0.001)	(0.001)	(0.001)
Parental income: greater than $100,000			0.006***	0.004***	0.003***
			(0.001)	(0.001)	(0.001)
High school fixed effects					✓
Number of families	2,626,286	2,626,286	2,626,286	2,626,286	2,626,286

Notes: The sample includes all families with 2 children. Regressions that including siblings' average PSAT scores, parental education, and parental income also include indicators for whether those variables are missing for the family.

TABLE A2—DESCRIPTIVE RELATIONSHIP BETWEEN SIBLING VOTING PATTERNS.

	(1)	(2)	(3)	(4)
Panel A: Outcome = younger sibling voted in 1st election				
Full sample, older sibling's 1st election is presidential				
Older sibling voted in 1st election	0.102***	0.102***	0.100***	0.092***
	(0.001)	(0.001)	(0.001)	(0.001)
Number of sibling pairs	1,284,032	1,284,032	1,284,032	1,281,545
Full sample, older sibling's 1st election is presidential				
Older sibling voted in younger sibling's 1st election	0.167***	0.167***	0.164***	0.158***
	(0.001)	(0.001)	(0.001)	(0.001)
Number of sibling pairs	1,284,032	1,284,032	1,284,032	1,281,545
Older sibling RD sample, 18th birthdays near any election				
Older sibling voted in election closest to 18th birthday	0.101***	0.101***	0.097***	0.089***
	(0.003)	(0.003)	(0.003)	(0.003)
Number of sibling pairs	202,931	202,931	202,931	199,269
Older sibling RD sample, 18th birthdays near presidential election				
Older sibling voted in election closest to 18th birthday	0.097***	0.097***	0.094***	0.086***
	(0.004)	(0.004)	(0.004)	(0.004)
Number of sibling pairs	112,770	112,770	112,770	108,893
Panel B: Outcome = older sibling voted in younger sibling's 1st election				
Full sample				
Younger sibling voted in 1st election	0.189***	0.189***	0.186***	0.178***
	(0.001)	(0.001)	(0.001)	(0.001)
Number of sibling pairs	2,575,240	2,575,240	2,575,240	2,573,041
Full sample, younger sibling's 1st election is presidential				
Younger sibling voted in 1st election	0.183***	0.183***	0.180***	0.168***
	(0.001)	(0.001)	(0.001)	(0.001)
Number of sibling pairs	1,343,926	1,343,926	1,343,926	1,341,584
Younger sibling RD sample, 18th birthdays near any election				
Younger sibling voted in election closest to 18th birthday	0.149***	0.149***	0.144***	0.138***
	(0.004)	(0.004)	(0.004)	(0.004)
Number of sibling pairs	165,734	165,734	165,734	161,876
Younger sibling RD sample, 18th birthdays near presidential election				
Younger sibling voted in election closest to 18th birthday	0.151***	0.151***	0.145***	0.140***
	(0.004)	(0.004)	(0.004)	(0.005)
Number of sibling pairs	90,997	90,997	90,996	86,857
Election year by state fixed effects	✓	✓	✓	✓
Sibling demographics		✓	✓	✓
Sibling PSAT scores			✓	✓
Younger sibling high school fixed effects				✓

Notes: The samples includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months. Robust standard errors are reported.

TABLE A3—RD SAMPLE COVARIATE BALANCE CHECKS.

	Sample		
	Own-own all sibling RD sample (1)	Older sibling RD sample (2)	Younger sibling RD sample (3)
Own characteristics			
Female	0.006	0.001	0.005
	(0.004)	(0.006)	(0.007)
PSAT score	0.278*	0.163	0.300
	(0.166)	(0.268)	(0.294)
White	-0.002	0.004	-0.008
	(0.003)	(0.006)	(0.006)
Black	-0.002	-0.006*	-0.002
	(0.002)	(0.003)	(0.004)
Asian	0.002	-0.002	0.006*
	(0.002)	(0.003)	(0.004)
Hispanic	0.003	0.007*	0.002
	(0.003)	(0.004)	(0.004)
Sibling characteristics			
Female		0.004	-0.003
		(0.006)	(0.007)
PSAT score		0.053	0.053
		(0.266)	(0.298)
White		0.004	-0.011*
		(0.006)	(0.006)
Black		-0.006*	-0.002
		(0.003)	(0.004)
Asian		-0.001	0.006*
		(0.003)	(0.004)
Hispanic		0.005	0.003
		(0.004)	(0.004)
Sibling age difference (months)		-0.062	0.121
		(0.080)	(0.129)
Age when 1st eligible to vote		0.003	-0.005
		(0.005)	(0.007)

Notes: This table reports covariance balance tests for our three main RD samples. To carry out these tests, we use the same reduced form RD specifications while using individual or sibling characteristics as the outcome. Each coefficient comes from a separate regression. For the older sibling RD sample, the sibling characteristics correspond to the younger sibling. Likewise, for the younger sibling RD sample, the sibling characteristics correspond to the older sibling. The sample in column 1 includes all siblings with an 18th birthday within 30 days of a presidential election. The sample in column 2 includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months and where the older sibling turned 18 near a presidential election. The sample includes sibling pairs from families with 2 children with age differences between 9 and 48 months and where the younger sibling turned 18 near a presidential election. In all cases, bandwidths are 30 days. Robust standard errors are reported in parentheses.

TABLE A4—ROBUSTNESS CHECKS FOR OWN FUTURE VOTING RD, ALL SIBLING SAMPLE.

	BW = 15 days (1)	BW = 30 days (2)	BW = 30 days; triangle weight (3)	BW = 30 days; quadratic (4)	BW = 60 days (5)	Optimal BW (6)
Outcome = Voted in election 2 years later						
Eligible to vote at age 18	0.020***	0.021***	0.020***	0.020***	0.021***	0.022***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
						{52}
Observations	144,416	286,614	286,614	286,614	584,875	507,346
Outcome = Voted in election 4 years later						
Eligible to vote at age 18	0.015***	0.016***	0.015***	0.015***	0.017***	0.017***
	(0.005)	(0.003)	(0.005)	(0.005)	(0.002)	(0.003)
						{55}
Observations	144,416	286,614	286,614	286,614	584,875	548,759
Outcome = Ever voted in the future						
Eligible to vote at age 18	0.016***	0.014***	0.014***	0.017***	0.016***	0.015***
	(0.005)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
						{53}
Observations	144,416	286,614	286,614	286,614	584,875	528,471
Outcome = Number of future general election votes						
Eligible to vote at age 18	0.059***	0.057***	0.063***	0.065***	0.055***	0.056***
	(0.013)	(0.009)	(0.013)	(0.014)	(0.006)	(0.007)
						{56}
Observations	144,416	286,614	286,614	286,614	584,875	548,759
Outcome = Number of future primary election votes						
Eligible to vote at age 18	0.016***	0.022***	0.014**	0.017***	0.024***	0.024***
	(0.006)	(0.004)	(0.006)	(0.006)	(0.003)	(0.003)
						{67}
Observations	144,416	286,614	286,614	286,614	584,875	651,477

Notes: The sample includes all siblings with an 18th birthday within 30 days of a presidential election. Regressions include fixed effects for the election year that is closest to their 18th birthday. Column 3 includes a triangular weight. Column 4 includes second order terms of the running variable on each side of the threshold. Column 6 uses the 'rdrobust' package in Stata to compute estimates with the 'optimal' bandwidths, which are reported in curly brackets. Robust standard errors are reported in parentheses.

TABLE A5—EFFECT OF ELIGIBILITY AT AGE 18 ON OWN FUTURE VOTING, FULL SAMPLE.

	First stage	Own future voting outcomes				
	Voted at age 18 (1)	Voted in election 2 years later (2)	Voted in election 4 years later (3)	Ever voted in the future (4)	General election votes (5)	Primary election votes (6)
Panel A: 18th birthdays near any election						
Eligible to vote at age 18	0.178***	0.014***	0.009***	0.014***	0.042***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)
Outcome control mean	0.005	0.144	0.201	0.412	0.798	0.148
Observations	1,864,816	1,864,816	1,864,816	1,864,816	1,864,816	1,864,816
Panel B: 18th birthdays near presidential elections						
Eligible to vote at age 18	0.253***	0.018***	0.015***	0.018***	0.053***	0.018***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.002)
Outcome control mean	0.006	0.071	0.276	0.399	0.753	0.143
Observations	1,098,813	1,098,813	1,098,813	1,098,813	1,098,813	1,098,813
Panel C: 18th birthdays near midterm elections						
Eligible to vote at age 18	0.069***	0.008***	-0.000	0.008***	0.025***	0.004
	(0.001)	(0.002)	(0.001)	(0.002)	(0.006)	(0.002)
Outcome control mean	0.005	0.250	0.092	0.431	0.861	0.155
Observations	766,003	766,003	766,003	766,003	766,003	766,003

Notes: The sample includes all individuals with an 18th birthday within 30 days of an election. Regressions include fixed effects for the election year that is closest to their 18th birthday. Bandwidth is 30 days. Robust standard errors are reported.

TABLE A6—EFFECT OF ELIGIBILITY AT AGE 18 ON OWN FUTURE VOTING, OLDER SIBLINGS SAMPLE.

	First stage	Own future voting outcomes				
	Voted at age 18 (1)	Voted in election 2 years later (2)	Voted in election 4 years later (3)	Ever voted in the future (4)	General election votes (5)	Primary election votes (6)
Panel A: 18th birthdays near any election						
Eligible to vote at age 18	0.207***	0.015***	0.012***	0.011**	0.054***	0.017***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.013)	(0.006)
Outcome control mean	0.008	0.183	0.213	0.512	1.134	0.218
Observations	202,937	202,937	202,937	202,937	202,937	202,937
Panel B: 18th birthdays near presidential elections						
Eligible to vote at age 18	0.295***	0.027***	0.018***	0.014**	0.074***	0.026***
	(0.004)	(0.004)	(0.005)	(0.006)	(0.018)	(0.008)
Outcome control mean	0.009	0.090	0.288	0.506	1.149	0.224
Observations	112,778	112,778	112,778	112,778	112,778	112,778
Panel C: 18th birthdays near midterm elections						
Eligible to vote at age 18	0.097***	0.001	0.003	0.007	0.029	0.005
	(0.003)	(0.006)	(0.004)	(0.007)	(0.018)	(0.008)
Outcome control mean	0.007	0.300	0.120	0.521	1.114	0.210
Observations	90,159	90,159	90,159	90,159	90,159	90,159

Notes: The sample includes the older siblings from sibling pairs that include the oldest child in the family with age differences between 9 and 48 months. Regressions include fixed effects for the election year that is closest to the older sibling's 18th birthday. Bandwidth is 30 days. Robust standard errors are reported.

TABLE A7—ROBUSTNESS CHECKS FOR OLDER SIBLING RD.

	BW = 15 days (1)	BW = 30 days (2)	BW = 30 days; triangle weight (3)	BW = 30 days; quadratic (4)	BW = 60 days (5)	Optimal BW (6)
Outcome = Younger sibling voted in 1st election						
Eligible to vote at age 18	0.011	0.010**	0.012*	0.011	0.006*	0.011**
	(0.007)	(0.005)	(0.007)	(0.007)	(0.003)	(0.005)
						{34}
Observations	56,551	112,778	112,778	112,778	230,094	127,004
Outcome = Younger sibling voted in 1st presidential election						
Eligible to vote at age 18	0.010	0.011**	0.014*	0.012	0.005	0.008*
	(0.008)	(0.005)	(0.008)	(0.008)	(0.004)	(0.005)
						{41}
Observations	56,551	112,778	112,778	112,778	230,094	158,745
Outcome = Younger sibling voted in 1st midterm election						
Eligible to vote at age 18	-0.003	-0.001	-0.002	-0.005	-0.001	0.000
	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)
						{39}
Observations	56,548	112,772	112,772	112,772	230,083	146,375
Outcome = Younger sibling ever voted						
Eligible to vote at age 18	-0.002	0.004	-0.000	-0.001	0.005	0.009*
	(0.008)	(0.006)	(0.008)	(0.009)	(0.004)	(0.005)
						{47}
Observations	56,551	112,778	112,778	112,778	230,094	183,072
Outcome = Younger sibling's number of votes cast						
Eligible to vote at age 18	0.019	0.042**	0.017	0.024	0.017	0.019
	(0.029)	(0.020)	(0.029)	(0.030)	(0.014)	(0.014)
						{61}
Observations	56,551	112,778	112,778	112,778	230,094	232,286

Notes: The sample includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months and where the older sibling turned 18 near the 2004, 2008, or 2012 elections. Regressions include fixed effects for the election year the younger sibling is first eligible for. Column 3 includes a triangular weight. Column 4 includes second order terms of the running variable on each side of the threshold. Column 6 uses the 'rdrobust' package in Stata to compute estimates with the 'optimal' bandwidths, which are reported in curly brackets. Robust standard errors are reported in parentheses.

TABLE A8—OLDER SIBLING RD PLACEBO CHECK, 18TH BIRTHDAYS NEAR ODD YEAR SPECIAL ELECTION DATES.

	Younger sibling voting outcomes						
	Voted in 1st election (1)	Voted in 1st pres. election (2)	Voted in 1st midterm election (3)	Ever voted (4)	Vote count (5)	General election vote count (6)	Primary election vote count (7)
Older sibling 'eligible' at age 18 in odd year	-0.005	-0.002	-0.004	-0.001	-0.009	-0.008	-0.002
	(0.004)	(0.004)	(0.003)	(0.005)	(0.014)	(0.011)	(0.005)
Outcome control mean	0.204	0.306	0.096	0.482	1.053	0.889	0.164
Observations	173,803	173,803	173,797	173,803	173,803	173,803	173,803

Notes: The sample includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months. The running variable in these RD specifications are defined as the number of days between the older sibling's 18th birthday and the nearest odd year election date, using the 'special election' dates in those years (2005, 2007, 2009, 2011). Regressions include fixed effects for the election year the younger sibling is first eligible for. Bandwidth is 30 days. Robust standard errors are reported.

TABLE A9—OLDER SIBLING RD, 18TH BIRTHDAYS NEAR ANY ELECTION.

	First stage	Younger sibling voting outcomes						
	Older sibling voted at age 18 (1)	Voted in 1st election (2)	Voted in 1st pres. election (3)	Voted in 1st midterm election (4)	Ever voted (5)	Vote count (6)	General election vote count (7)	Primary election vote count (8)
Older sibling eligible	0.206***	0.001	0.005	-0.003	-0.002	0.011	0.008	0.003
to vote at age 18 (RF)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.014)	(0.011)	(0.005)
	[6,262]							
Older sibling voted		0.003	0.022	-0.013	-0.009	0.054	0.040	0.014
at age 18 (IV)		(0.017)	(0.020)	(0.013)	(0.021)	(0.068)	(0.054)	(0.023)
Outcome control mean	0.008	0.209	0.304	0.097	0.493	1.142	0.968	0.174
Observations	202,937	202,937	202,937	202,931	202,937	202,937	202,937	202,937

Notes: The sample includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months and where the older sibling turned 18 near the 2004, 2006, 2008, 2010, or 2012 elections. Regressions include fixed effects for the election year the younger sibling is first eligible for. Bandwidth is 30 days. First stage F-stat reported in brackets in column 1. Robust standard errors are reported in parentheses.

TABLE A10—OLDER SIBLING RD BY SIBLING SEX COMPOSITION.

	1st stage	Younger sibling voting outcomes						
	Older sibling voted at age 18 (1)	Voted in 1st election (2)	Voted in 1st pres. election (3)	Voted in 1st midterm election (4)	Ever voted (5)	Vote count (6)	General election vote count (7)	Primary election vote count (8)
Panel A: Sisters								
Older sibling eligible to vote at age 18 (RF)	0.277*** (0.007) [1,485]	0.029*** (0.009)	0.039*** (0.010)	0.003 (0.007)	0.031*** (0.011)	0.140*** (0.037)	0.099*** (0.030)	0.041*** (0.012)
Older sibling voted at age 18 (IV)		0.105*** (0.033)	0.141*** (0.037)	0.012 (0.024)	0.111*** (0.041)	0.506*** (0.131)	0.356*** (0.107)	0.150*** (0.042)
Outcome control mean	0.004	0.220	0.291	0.092	0.486	1.173	1.003	0.170
Observations	31,014	31,014	31,014	31,012	31,014	31,014	31,014	31,014
Panel B: Brothers								
Older sibling eligible to vote at age 18 (RF)	0.321*** (0.008) [1,457]	0.014 (0.010)	0.013 (0.011)	-0.003 (0.007)	-0.008 (0.012)	0.026 (0.041)	0.027 (0.032)	-0.001 (0.014)
Older sibling voted at age 18 (IV)		0.045 (0.031)	0.041 (0.035)	-0.009 (0.023)	-0.024 (0.038)	0.082 (0.128)	0.084 (0.101)	-0.002 (0.045)
Outcome control mean	0.013	0.227	0.282	0.101	0.509	1.237	1.041	0.196
Observations	26,638	26,638	26,638	26,636	26,638	26,638	26,638	26,638
Panel C: Older sibing female, younger sibling male								
Older sibling eligible to vote at age 18 (RF)	0.276*** (0.008) [1,296]	-0.002 (0.010)	0.004 (0.011)	-0.003 (0.008)	0.000 (0.012)	-0.004 (0.043)	0.001 (0.033)	-0.005 (0.015)
Older sibling voted at age 18 (IV)		-0.006 (0.036)	0.015 (0.040)	-0.010 (0.027)	0.001 (0.044)	-0.016 (0.154)	0.003 (0.119)	-0.019 (0.054)
Outcome control mean	0.004	0.238	0.294	0.109	0.515	1.298	1.086	0.211
Observations	27,566	27,566	27,566	27,566	27,566	27,566	27,566	27,566
Panel D: Older sibing male, younger sibling female								
Older sibling eligible to vote at age 18 (RF)	0.309*** (0.008) [1,445]	-0.003 (0.010)	-0.014 (0.011)	-0.003 (0.007)	-0.009 (0.012)	-0.002 (0.040)	-0.003 (0.032)	0.001 (0.013)
Older sibling voted at age 18 (IV)		-0.011 (0.032)	-0.046 (0.036)	-0.008 (0.023)	-0.028 (0.039)	-0.008 (0.129)	-0.009 (0.103)	0.002 (0.043)
Outcome control mean	0.015	0.232	0.303	0.096	0.494	1.209	1.031	0.178
Observations	27,558	27,558	27,558	27,558	27,558	27,558	27,558	27,558

Notes: The sample includes sibling pairs that include the oldest child in the family with age differences between 9 and 48 months and where the older sibling turned 18 near the 2004, 2008, or 2012 elections. Regressions include fixed effects for the election year the younger sibling is first eligible for. Bandwidth is 30 days. First stage F-stat reported in brackets in column 1. Robust standard errors are reported in parentheses.

TABLE A11—ROBUSTNESS CHECKS FOR YOUNGER SIBLING RD.

	BW = 15 days (1)	BW = 30 days (2)	BW = 30 days; triangle weight (3)	BW = 30 days; quadratic (4)	BW = 30 days; donut (5)	BW = 60 days (6)	Optimal BW (7)
Younger sibling's 18th birthday near presidential election							
Eligible to vote at age 18	0.021**	0.013**	0.027***	0.030***	0.011*	0.012***	0.012***
	(0.009)	(0.006)	(0.009)	(0.009)	(0.006)	(0.004)	(0.005)
							{52}
Observations	45,930	90,997	90,997	90,997	89,901	185,525	162,233
Younger sibling's 18th birthday near any election							
Eligible to vote at age 18	0.015***	0.009**	0.018***	0.018***	0.007*	0.008***	0.008**
	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.003)	(0.003)
							{46}
Observations	84,069	165,751	165,751	165,751	163,276	336,657	255,744

Notes: The sample includes sibling pairs from families with 2 children with age differences between 9 and 48 months and where the younger sibling turned 18 near the 2004, 2008, or 2012 elections. Regressions include fixed effects for the election year the younger sibling is first eligible for. Column 3 includes a triangular weight. Column 4 includes second order terms of the running variable on each side of the threshold. Column 6 uses the 'rdrobust' package in Stata to compute estimates with the 'optimal' bandwidths, which are reported in curly brackets. Robust standard errors are reported in parentheses.

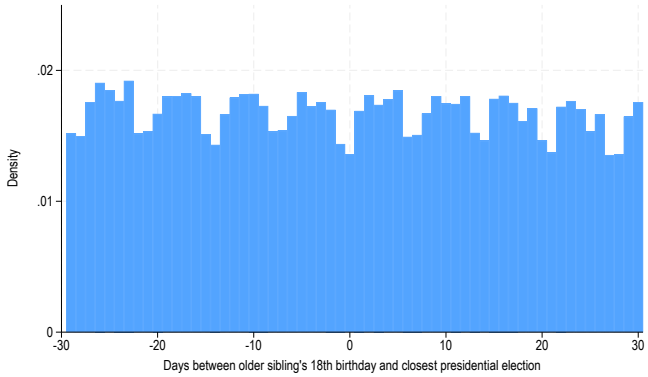
TABLE A12—YOUNGER SIBLING RD BY SIBLING SEX COMPOSITION.

	Younger sibling voted in election closest to 18th birthday	Older sibling voted in election closest to younger sibling's 18th birthday	
	First stage (1)	Reduced Form (2)	Second Stage (3)
Panel A: Sisters			
RD estimate	0.339***	0.006	0.017
	(0.009)	(0.011)	(0.034)
First stage F-stat	[1,581]		
Outcome control mean	0.003	0.289	0.289
Observations	24,577	24,577	24,577
Panel B: Brothers			
RD estimate	0.317***	0.011	0.033
	(0.009)	(0.012)	(0.039)
First stage F-stat	[1,235]		
Outcome control mean	0.003	0.291	0.291
Observations	21,384	21,384	21,384
Panel C: Older sibling female, younger sibling male			
RD estimate	0.327***	0.033***	0.100***
	(0.009)	(0.012)	(0.037)
First stage F-stat	[1,281]		
Outcome control mean	0.012	0.296	0.296
Observations	22,414	22,414	22,414
Panel D: Older sibling male, younger sibling female			
RD estimate	0.344***	0.005	0.015
	(0.009)	(0.012)	(0.035)
First stage F-stat	[1,477]		
Outcome control mean	0.005	0.293	0.293
Observations	22,622	22,622	22,622

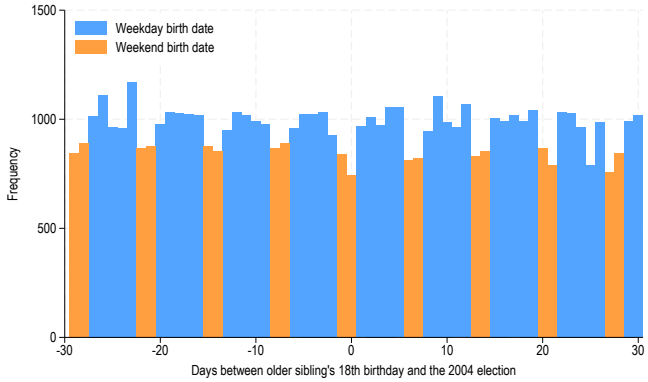
Notes: The sample includes sibling pairs from families with 2 children with age differences between 9 and 48 months and where the younger sibling turned 18 near the 2004, 2008, or 2012 elections. Regressions include fixed effects for the election year that is closest to the younger sibling's 18th birthday. Bandwidth is 30 days. Robust standard errors are reported.

FIGURE A1. DENSITY OF OLDER SIBLING'S 18TH BIRTHDAY RELATIVE TO PRESIDENTIAL ELECTIONS

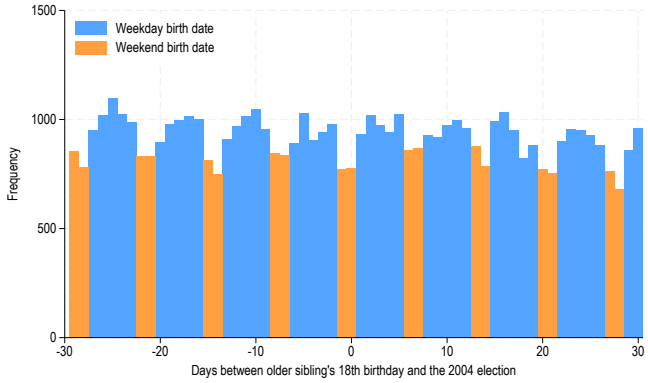
(a) All presidential elections



(b) 2004 election, 1986 birth year



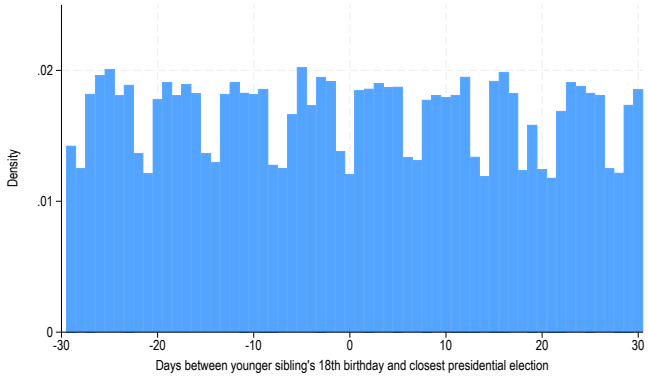
(c) 2008 election, 1990 birth year



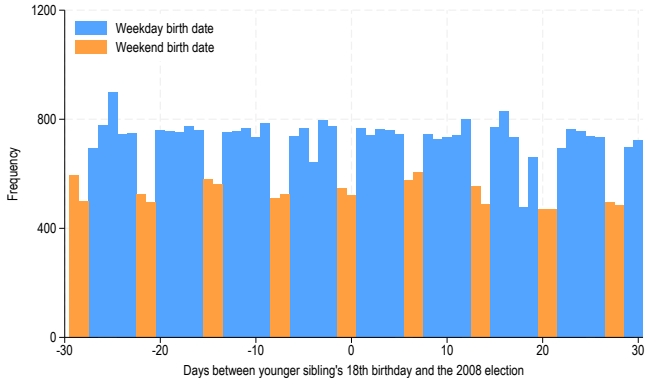
Note: This figure shows histograms of the running variable for the older sibling RD design. Panel (a) shows the histogram for all observations. Panels (b) and (c) show the histograms specifically for the 2004 and 2008 elections to show how the lumpiness of the densities are explained by fewer births on weekends.

FIGURE A2. DENSITY OF YOUNGER SIBLING'S 18TH BIRTHDAY RELATIVE TO PRESIDENTIAL ELECTIONS

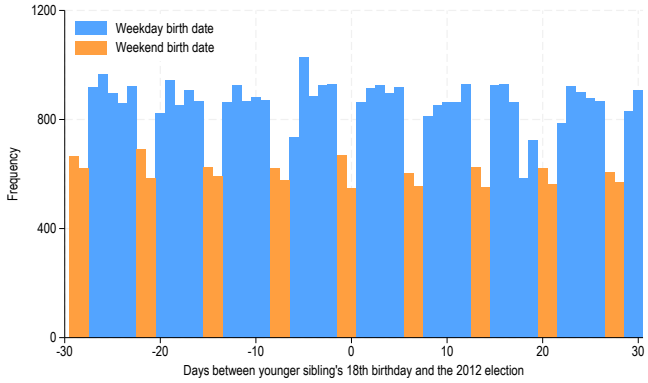
(a) All presidential elections



(b) 2008 election, 1990 birth year



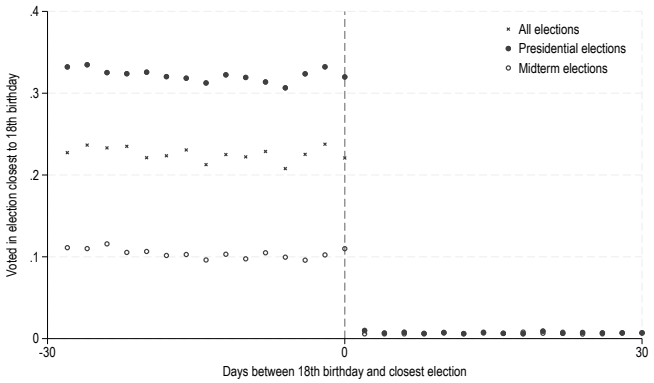
(c) 2012 election, 1994 birth year



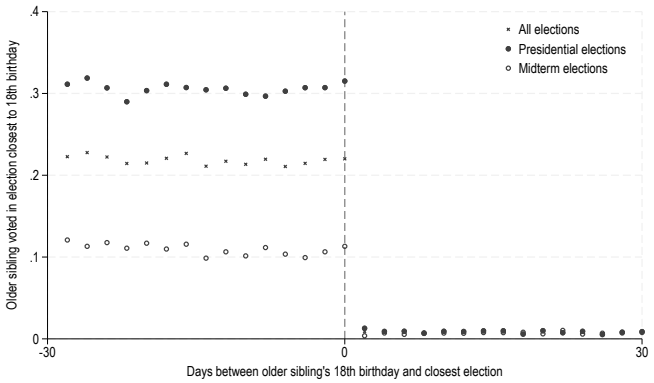
Note: This figure shows histograms of the running variable for the younger sibling RD design. Panel (a) shows the histogram for all observations. Panels (b) and (c) show the histograms specifically for the 2008 and 2012 elections to show how the lumpiness of the densities are explained by fewer births on weekends.

FIGURE A3. FIRST STAGE

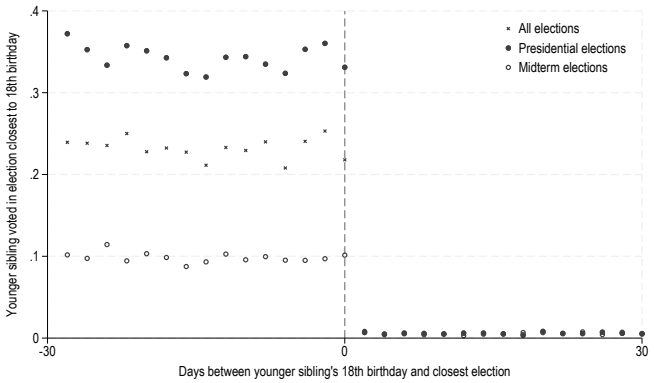
(a) All siblings RD sample



(b) Older sibling RD sample



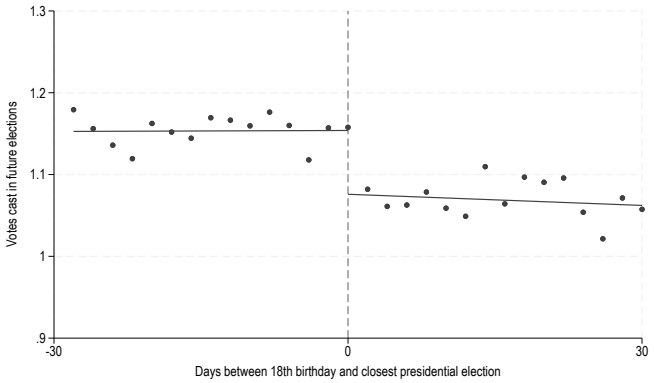
(c) Younger sibling RD sample



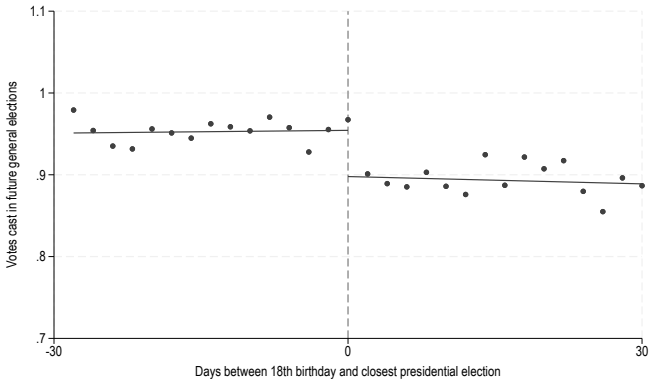
Note: This figure shows binned scatterplots of voting rates in the election closest to 18th birthdays by days between that birthday and the presidential elections of 2004, 2008, and 2012 for our main RD samples. Each bin consists of 2 days.

FIGURE A4. THE EFFECT OF AGE 18 ELIGIBILITY ON FUTURE VOTING - ADDITIONAL OUTCOMES

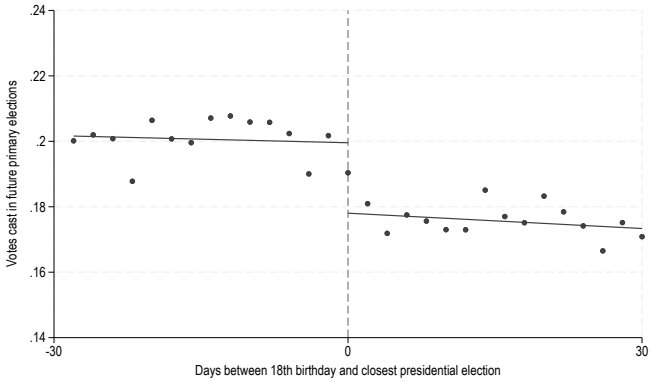
(a) Votes cast in future elections



(b) Votes cast in future general elections



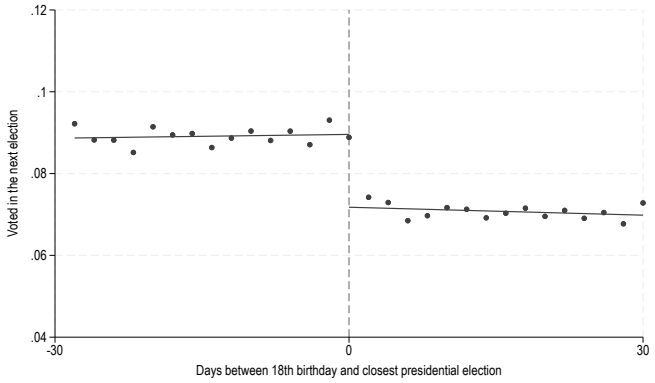
(c) Votes cast in future primary elections



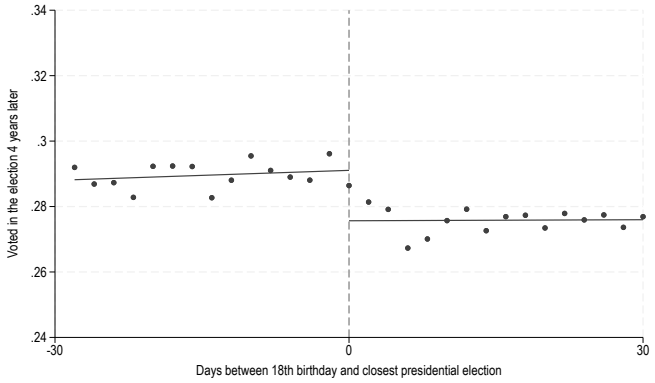
Note: This figure shows binned scatterplots of average voting outcomes by days between each sibling's 18th birthday and the presidential elections of 2004, 2008, and 2012. Each bin consists of 2 days.

FIGURE A5. THE EFFECT OF AGE 18 ELIGIBILITY ON FUTURE VOTING - FULL SAMPLE

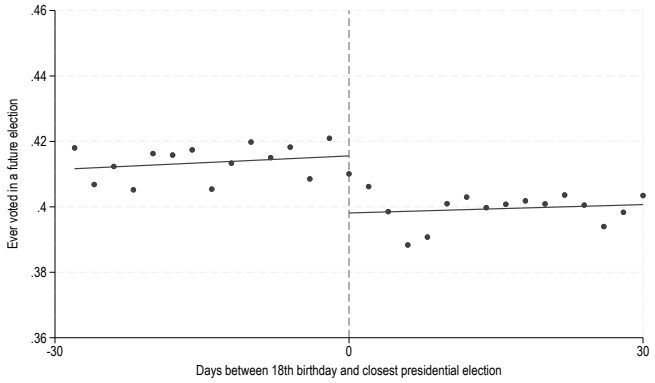
(a) Voted in election 2 years later



(b) Voted in election 4 years later



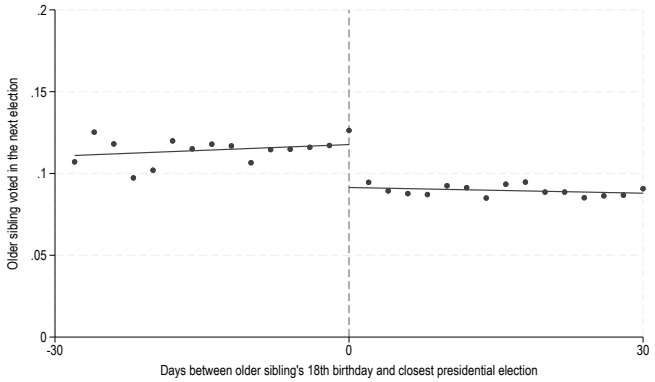
(c) Ever voted in future election



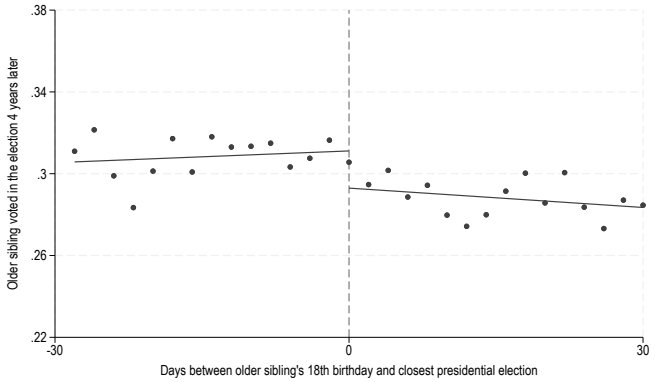
Note: This figure shows binned scatterplots of average voting outcomes by days between an individual's 18th birthday and the presidential elections of 2004, 2008, and 2012. Each bin consists of 2 days.

FIGURE A6. THE EFFECT OF AGE 18 ELIGIBILITY ON FUTURE VOTING - OLDER SIBLINGS

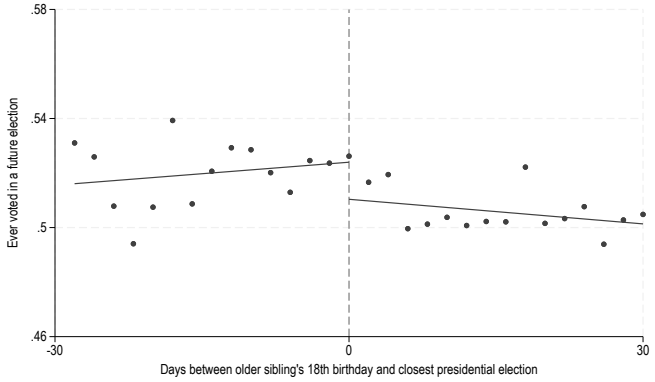
(a) Voted in election 2 years later



(b) Voted in election 4 years later



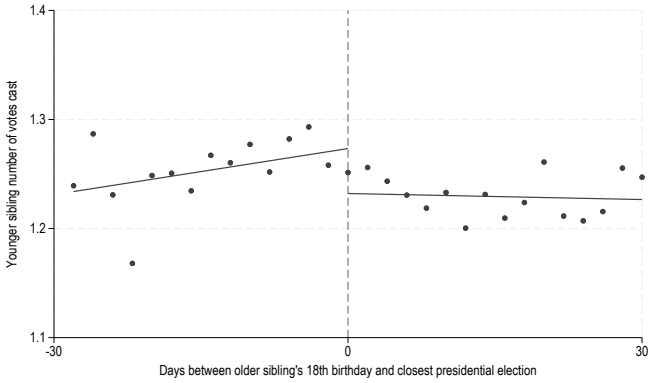
(c) Ever voted in future election



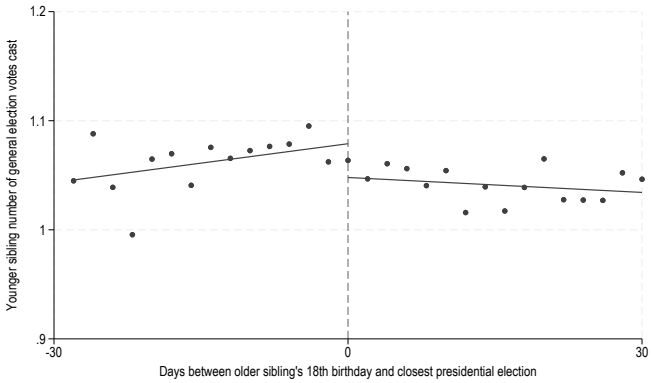
Note: This figure shows binned scatterplots of average older sibling voting outcomes by days between the older sibling's 18th birthday and the presidential elections of 2004, 2008, and 2012. Each bin consists of 2 days.

FIGURE A7. EFFECT OF OLDER SIBLING'S AGE 18 ELIGIBILITY ON YOUNGER SIBLING'S VOTING - ADDITIONAL OUTCOMES

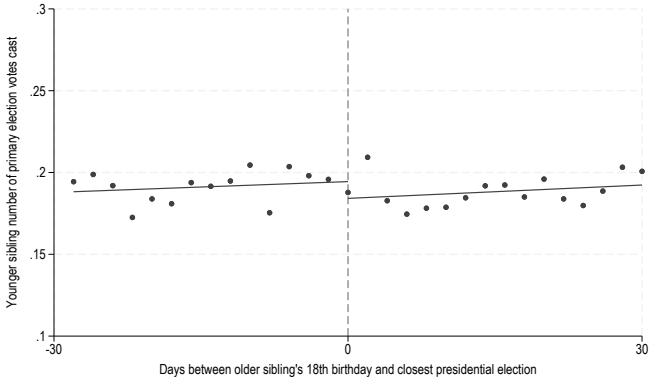
(a) Younger sibling votes cast



(b) Younger sibling votes cast in general elections



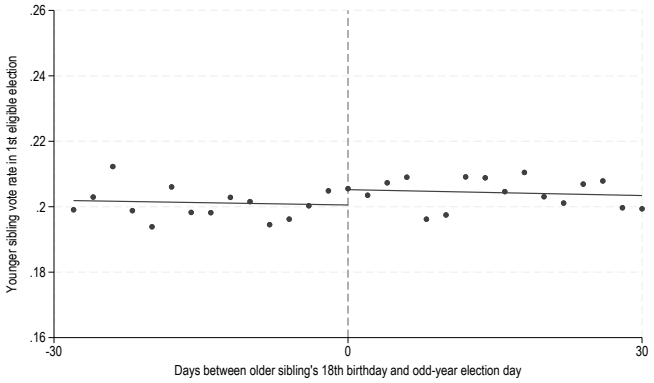
(c) Younger sibling votes cast in primary elections



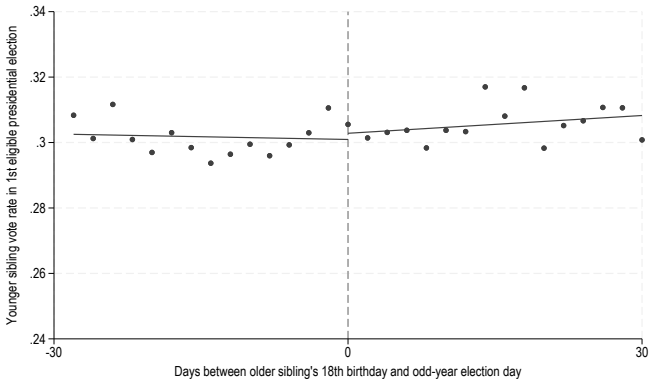
Note: This figure shows binned scatterplots of average younger sibling voting outcomes by days between the older sibling's 18th birthday and the presidential elections of 2004, 2008, and 2012. Voting outcomes are residualized for fixed effects of the younger sibling's first election to match our estimating equation. Each bin consists of 2 days.

FIGURE A8. PLACEBO TEST - EFFECT OF OLDER SIBLING'S AGE 18 ELIGIBILITY FOR 'ODD YEAR' ELECTIONS ON YOUNGER SIBLING'S VOTING

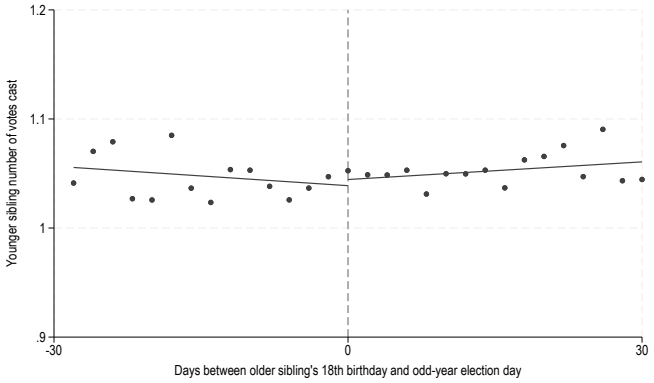
(a) Younger sibling voted in 1st election



(b) Younger sibling voted in 1st presidential election



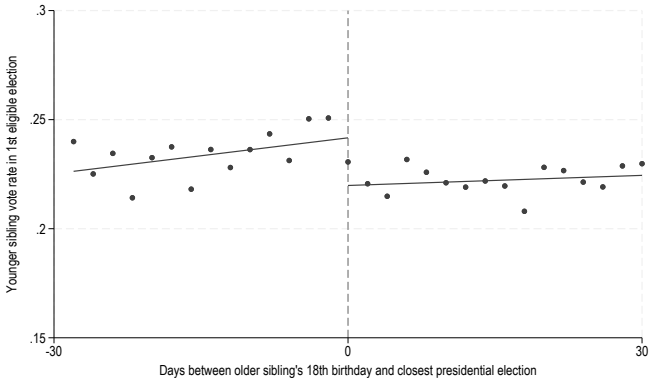
(c) Younger sibling votes cast



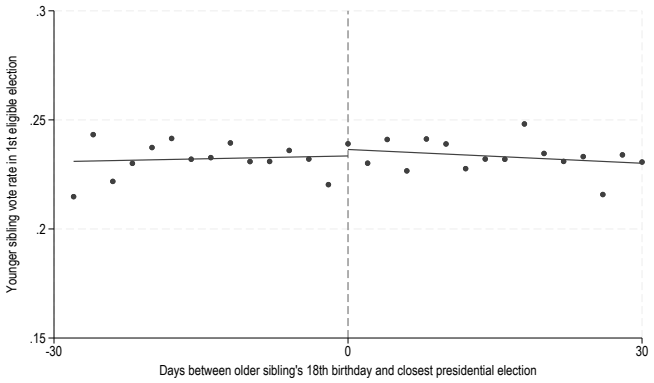
Note: This figure shows binned scatterplots of average younger sibling voting outcomes by days between the older sibling's 18th birthday and the nearest odd year election date, using the 'special election' dates in those years (2005, 2007, 2009, 2011). Each bin consists of 2 days.

FIGURE A9. EFFECT OF OLDER SIBLING'S AGE 18 ELIGIBILITY ON YOUNGER SIBLING'S VOTING - BY SIBLING SEX COMPOSITION

(a) Same sex sibling pairs



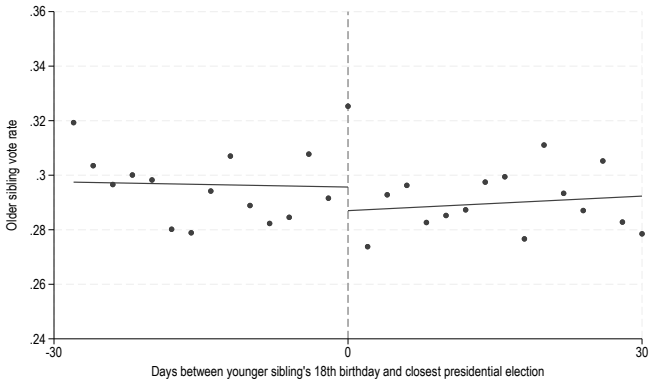
(b) Opposite sex sibling pairs



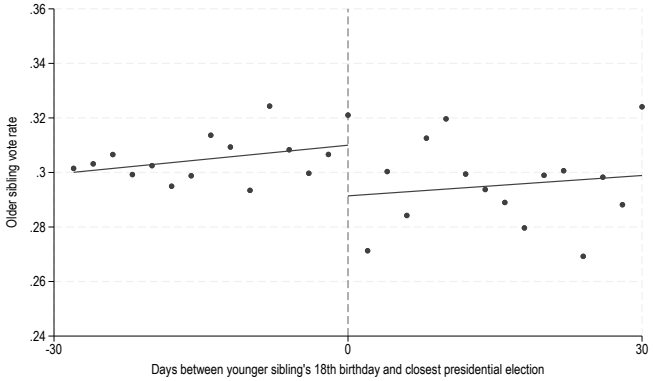
Note: This figure shows binned scatterplots of average younger sibling 1st election vote rates by days between the older sibling's 18th birthday and the presidential elections of 2004, 2008, and 2012. Voting outcomes are residualized for fixed effects of the younger sibling's first election to match our estimating equation. Each bin consists of 2 days. Panel A includes same sex sibling pairs (brother and sister pairs). Panel B includes opposite sex sibling pairs.

FIGURE A10. EFFECT OF YOUNGER SIBLING'S AGE 18 ELIGIBILITY ON OLDER SIBLING'S VOTING - BY SIBLING SEX COMPOSITION

(a) Same sex sibling pairs



(b) Opposite sex sibling pairs



Note: This figure shows binned scatterplots of average older sibling vote rates in their younger sibling's 1st election by days between the younger sibling's 18th birthday and the presidential elections of 2004, 2008, and 2012. Voting outcomes are residualized for fixed effects of the election closest to the younger sibling's 18th birthday to match our estimating equation. Each bin consists of 2 days. Panel A includes same sex sibling pairs (brother and sister pairs). Panel B includes opposite sex sibling pairs.

APPENDIX B - MATCHING PROCESS

In this appendix, we discuss the process of matching the education records with the voting records and then we assess match quality by comparing the matched data against a nationally representative dataset on voting rates. Since this study employs the same base data as Bell et al. (2022), material in this appendix is largely reproduced from Appendices A and B of Bell et al. (2022) with the authors' permission.

B1. Preparing the Education Data

The College Board data initially consists of 39.6 million observations across cohorts graduating high school between 2004 and 2018. We then restrict the sample to students who attended high school in one of the 50 states or Washington, DC, and those who have complete information on birth date (i.e. month, day, and year). This leaves over 37.6 million observations that serve as the base sample. Our analyses focus on the 21.3 million observations in the 2004 to 2012 cohorts.

B2. Preparing the Voting Data

We begin with a dataset that keeps voters with a birth year after 1984 to be consistent with the education data. We also keep voters where their year of birth is missing, leaving us with 52,921,628 observations.

With this initial voting dataset, the goal is to get one observation per person. Getting one observation per person presents two challenges relative to the education data. First, the DOB may be partially incomplete and in rare cases, entirely missing.[9] In the education data, we removed those observations from the base sample. We cannot drop those voting records because someone with missing DOB may be a match to the education data. The supposed match in the education data will look as if they never voted if we remove those voting observations. Second, the same person can show up more than once across states if they move (and potentially within the same state if a state's records are inexact and/or missing a piece of the birth date).

To address the above issues, we first go from one observation per person per state to one observation per person across states. This involves finding exact matches and then fuzzy matches using name and DOB, all while making sure there is no overlap in the same election, which is indicative of this being more than one person. We also create a series of flags, which we use in robustness tests, when there are common names or missing information on the DOB.

B3. Identifying Voters That Are Not Duplicates

We first identify all voter observations across all states that do not need to be deduplicated and set them aside to what will be the final dataset of deduplicated voters. To do so, we start with all observations with a unique first and last name combination across

[9] 4.5 percent had a missing birth year, 14.6 percent had a missing birth month, and 21 percent had a missing birth day.

all states. At the end of this step, 34.4 percent of the voter data are considered unique observations.

With the set of voters with non-unique first and last name combinations, we identify all observations that were unique by first name, last name, and birth year. After this step an additional 21.8 percent of the voter data are considered unique observations. We then identify all observations that were unique by first name, last name, birth year and birth month. At the end of this step, an additional 20.4 percent of the voter data are considered unique observations. Then we do the same by adding birth day, producing an additional 10.9 percent of the voter data considered unique. We do the same thing again with the addition of middle initial. After this step an additional 3.8 percent of the voter data are considered unique. Along each step, we remove the "unique" voter but generate a flag that describes how the uniqueness was determined.

At the end of this process, 91.3 percent of the voter data are considered unique observations.

B4. Identifying Duplicate Voters

With the remaining 8.7 percent of voting data, we identify the observations with the same DOB information, the same exact first and last name, but differing voting history (e.g. one observation voted in an election in 2010 but did not vote in 2012 and the observation with the same name and DOB information had the opposite voting pattern). The vast majority of these observations were collapsed into a single observation and we retained whether we collapsed two or three (Less than 1.1 percent of these observations were unable to be collapsed into a single observation). Almost all were initially two observations that we collapsed into one.

This deduplication process accounted for roughly 7.3 percent of the original voting observations, leaving just 1.5 percent of original voting observations. We take these observations and place them with the uniquely identified and deduplicated observations and treat them as unique.

From here, we created our final dataset to use in the matching to the College Board data. It contains each dataset that from each round of deduplication to identify unique observations and the remaining observations that were not uniquely identified. In all, we created a dataset with 50,975,728 observations of voting history.

After each de-duplication round, in which observations were uniquely identified, we created a flag to identify all the observations that we determined were uniquely identified for that round to use in a series of robustness tests.

B5. Exact Matching

We start by using exact matching of the education data to the voting data. An exact match required perfect agreement between the first name, last name, middle initial, and date of birth (DOB), none of which could be incomplete in either dataset for this step. We

identified 14,175,803 unique matches.[10] This implies that 35.5 percent of the education sample was found in the voting records through exact matching.

B6. Fuzzy Matching

After removing any observation that was exactly matched from both the education and voting datasets, we implement a fuzzy matching algorithm, in three broad steps.

First, we use College Board's 23 step fuzzy matching algorithm - a process they use in other applications. The algorithm starts by very slowly loosening the exact matching criteria. Specifically, it starts with exact matches on first name, last name, and DOB but one or both of the datasets are missing a middle initial, everything matches exactly but for one edit to one name, and everything matches but the first and last name are swapped. The most relaxed criteria - the 23rd step - matches exactly on first name, DOB, and gender but one of the two last names from the two datasets is a suffix of the other and only one of the middle initials is missing. After each step, the matched observations are not replaced for additional matches. We also retain each step in which the observations are matched for robustness tests. This process generates 6,547,716 additional matches, which is an additional 16.4 percent of the education observations.

Second, we used the data linkage method (also known as the editing distance method) employed by Dusetzina, Tyree, Meyer, Meyer, Green, & Carpenter (2014). The method calculates a probability that two strings are a match with the following formula:

$$\sum_{i=1}^{2}[1-(length(name_i)*spedis(name_i,name_{-i}))/2400)]/2 \tag{B1}$$

The two names from each dataset, indexed by i, are compared in both character length and also "spelling distance" (i.e. *spedis*). Spelling distance is a common function in statistical software that compares the letters in the name.[11]

We consider any value greater than 0.95 a high enough probability to be a match. We first do this for first names, maintaining an exact match on last name and DOB, and then again for last name. These generate an additional 27,557 matches (one percent of the education data).

Third and finally, we slightly loosen the criteria on birth date by using exact matches on everything previously described but the voting data has a missing birth day or birth month, but not both and not missing year, so there is no conflicting information. This generates 635,344 additional matches (nearly two percent of the education data).

In total, we matched approximately 27.7 million observations, accounting for almost 53.6 percent of the education sample.

[10] 83,652 students were matched to multiple voting records (usually two), creating 193,986 observations. In our main analyses, we randomly choose one of these matches but our results are entirely insensitive to alternatives analyses, largely because this impacted such a small fraction of the sample.

[11] Documentation for *spedis* in SAS, the statistical software we used, can be found here: https://support.sas.com/resources/papers/proceedings/proceedings/sugi25/25/cc/25p086.pdf.

B7. Assessing Match Quality

Next, we assess the quality of our match. To do so, we compare voting rates in our matched data for different groups and different elections to reported voting rates in the Voting and Registration dataset from the U.S. Census Bureau.[12] These data come from a supplement to the Current Population Survey. We focus on the voting rates among 18-24 year olds in the 2012 and 2016 elections, a time period that our matched data covers well.

In Table B1 through Table B3, we compare voting rates in the two datasets by election, age, race, sex, state, and sometimes combinations of those variables. Generally speaking, we find lower voting rates in our matched data than in the Census data, but the patterns across subgroups follow one another. The lower voting rates is expected, because our matching process is imperfect and somewhat conservative. And it is reassuring that the relative voting rates across subgroups generally match.

Table B1 shows voting rates in the two datasets by age and sex. The table shows that the voting increases with age (in both elections), according to the Census. Our matched data show a similar pattern, although at lower rates.

Table B1 also highlights the differences between the two samples by sex. Similar to the Census data, older males vote more than younger males in the matched data. However, this is not true for females, especially in the 2012 election. This is likely because we had difficulty matching to women who change their last name. This also explains why females vote at higher rates than males in both datasets for the youngest cohorts, before women typically get married and change their last name, but not so for older cohorts. For older cohorts, Census data suggest females are more likely to vote than males, but the matched data does not. These facts motivate some robustness tests that focus on males and recent cohorts.

Table B2 shows voting rates in the two datasets by race and sex, The table shows that Black people are the most likely to vote in 2012 in both elections and Asian people are least likely. Similarly, White people are most likely to vote in 2016 in both datasets and Asian people are the least likely. These similar patterns between the datasets is comforting.

Finally, voting rates in the two datasets are compared by state. It is immediately clear that our matched sample reflects the voting rates of the Census in some states better than others. This is partially because College Board has relatively low coverage in some states, like Mississippi. However, in states where College Board has substantial coverage, such as Virginia, the two voting rates are well aligned. This motivates a few additional robustness tests, including only using states where College Board has substantial coverage and only using states where the voting rates between these two datasets are well aligned.

[12] https://www.census.gov/topics/public-sector/voting.html

TABLE B1—COMPARISON OF VOTING RATES IN CENSUS DATA AND MATCHED DATA IN 2012 AND 2016 ELECTIONS, BY AGE AND SEX.

		Voted in 2012 (Percent)			Voted in 2016 (Percent)		
	Census Age (in years)	High School Cohort	Census	Matched Data	High School Cohort	Census	Matched Data
Overall	18	2012	31.3	25.2	2016	32.3	30.0
	19	2011	33.9	25.2	2015	36.9	30.9
	20	2010	39.7	24.7	2014	40.1	31.0
	21	2009	37.5	24.8	2013	38.8	31.5
	22	2008	41.0	26.9	2012	41.4	33.0
	23	2007	40.0	26.7	2011	41.6	33.1
	24	2006	41.5	26.6	2010	43.8	33.5
Male	18	2012	27.8	23.2	2016	30.5	27.1
	19	2011	30.4	22.8	2015	34.4	27.2
	20	2010	34.4	22.7	2014	37.4	27.3
	21	2009	34.3	23.3	2013	35.8	27.9
	22	2008	38.3	26.3	2012	36.3	29.7
	23	2007	39.3	27.1	2011	38.5	30.4
	24	2006	37.1	28.2	2010	41.9	31.6
Female	18	2012	34.8	27.2	2016	34.3	32.8
	19	2011	37.5	27.5	2015	39.5	34.4
	20	2010	44.7	26.5	2014	42.7	34.5
	21	2009	40.9	26.2	2013	42.0	35.0
	22	2008	43.7	27.5	2012	46.1	36.3
	23	2007	40.8	26.4	2011	44.8	35.8
	24	2006	46.0	25.1	2010	45.8	35.2

Notes: The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age.

TABLE B2—COMPARISON OF VOTING RATES IN CENSUS DATA AND MATCHED DATA IN 2012 AND 2016 ELECTIONS, BY RACE AND SEX.

	Voted in 2012 (Percent)		Voted in 2016 (Percent)	
	Census	Matched Data	Census	Matched Data
Total	38.0	25.7	39.4	31.8
Male	34.7	24.7	36.5	28.7
Female	41.3	26.7	42.4	34.9
Asian	20.1	17.2	25.2	25.9
Black	45.9	33.5	40.2	29.1
Hispanic	26.7	20.1	27.2	26.8
White	37.8	26.8	41.1	36.4
Male + Asian	18.1	15.3	21.2	21.8
Male + Black	41.4	26.5	36.5	22.1
Male + Hispanic	24.0	17.7	21.8	22.5
Male + White	34.6	28.0	38.3	34.9
Female + Asian	22.1	19.1	29.5	30.0
Female + Black	50.2	39.6	43.7	35.4
Female + Hispanic	29.7	22.3	32.7	30.7
Female + White	41.0	25.7	43.9	37.8

Notes: Census voting rates are for 18-24 year-olds; 2012 voting rates in matched data are for the 2006-2012 high school graduation cohorts; 2016 voting rates in matched data are for the 2010-2016 high school graduation cohorts. The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age.

TABLE B3—COMPARISON OF VOTING RATES IN CENSUS DATA AND MATCHED DATA IN 2012 AND 2016 ELECTIONS, BY STATE.

	Voted in 2012 (Percent)		Voted in 2016 (Percent)	
State	Census	Matched Data	Census	Matched Data
Mississippi	62.4	17.3	46.1	18.1
Minnesota	57.0	21.4	49.6	22.5
Wisconsin	53.3	24.7	45.6	21.2
Colorado	52.5	23.0	43.1	25.8
South Carolina	51.3	28.4	42.7	33.9
New Hampshire	50.0	17.8	*	21.1
Oregon	47.6	25.7	45.2	26.4
Iowa	46.6	41.1	35.5	52.9
Massachusetts	45.6	33.5	39.9	42.8
Rhode Island	45.2	27.2	*	35.9
North Carolina	45.1	16.8	44.8	18.5
Ohio	44.7	38.1	39.6	43.1
Michigan	43.5	17.1	36.1	22.3
Maine	42.8	13.7	48.8	14.3
Missouri	42.5	33.2	45.9	43.5
Maryland	42.1	33.3	48.0	40.1
Virginia	42.0	37.1	54.6	42.9
Montana	40.7	33.3	*	40.2
Delaware	40.5	21.0	*	22.0
Louisiana	40.4	11.6	49.2	14.9
Arizona	40.0	12.3	35.8	17.8
Pennsylvania	39.9	27.1	48.7	40.5
Nevada	38.2	27.3	37.6	32.3
Washington	38.2	33.7	42.2	38.7
Georgia	37.9	17.1	40.5	15.6
Kentucky	37.1	28.7	51.1	40.3
New Mexico	37.1	11.3	37.8	12.3
Connecticut	36.9	27.9	37.0	37.7
Florida	36.8	32.0	33.1	36.9
California	36.5	23.0	37.5	33.9
Nebraska	36.2	30.7	50.1	45.2
Indiana	35.7	24.8	40.8	31.9
Alabama	35.3	34.0	41.4	40.0
New York	35.0	27.1	34.6	33.2
New Jersey	34.9	32.4	35.9	38.4
Utah	34.8	28.2	41.6	38.3
Tennessee	34.0	26.9	29.9	32.7
South Dakota	32.7	12.3	*	16.6
Illinois	32.2	29.1	45.3	41.6
Kansas	30.1	26.5	33.8	36.1
Idaho	29.8	9.2	40.7	13.3
Oklahoma	27.2	21.9	32.4	34.2
Arkansas	24.3	26.0	33.1	30.9
West Virginia	22.6	28.1	32.2	38.0
Texas	22.5	19.1	27.3	24.1
Hawaii	22.1	8.0	20.4	10.1
District of Columbia	*	19.9	*	20.9
North Dakota	*	15.0	*	19.2
Alaska	*	12.4	*	15.1
Vermont	*	11.3	*	14.9
Wyoming	*	9.2	*	16.5

Notes: Census voting rates are for 18-24 year-olds; voting rates in matched data are for the 2006-2012 high school graduation cohorts. The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age. *indicates that Census data are not available.