Political Analysis Advance Access published August 18, 2009

doi:10.1093/pan/mpp014

Driving Saints to Sin: How Increasing the Difficulty of Voting Dissuades Even the Most Motivated Voters

John E. McNulty
Department of Political Science, Binghamton University, P.O. Box 6000 Binghamton, NY 13902-6000
e-mail: jmcnulty@binghamton.edu (corresponding author)

Conor M. Dowling
Institution for Social & Policy Studies, Yale University, 77 Prospect Street, New Haven, CT 06520 e-mail: conor.dowling@yale.edu

Margaret H. Ariotti
Department of Political Science and Department of Geography, Binghamton University, P.O. Box 6000 Binghamton, NY 13902-6000 e-mail: molly@binghamton.edu

The consolidation of polling places in the Vestal Central School District in New York State during the district's 2006 budget referendum provides a naturalistic setting to study the effects of polling consolidation on voter turnout on an electorate quite distinct from previous work by Brady and McNulty (2004, The costs of voting: Evidence from a natural experiment. Presented at the Annual Meeting of the Society for Political Methodology, Palo Alto, CA). In particular, voters in local elections are highly motivated and therefore might be thought to be less affected by poll consolidation. Nevertheless, through a matching analysis we find that polling consolidation decreases voter turnout substantially, by about seven percentage points, even among this electorate, suggesting that even habitual voters can be dissuaded from going to the polls. This finding has implications for how election administrators ought to handle cost-cutting measures like consolidation.

The questions of who votes and why have occupied several generations of political scientists (e.g., Merriam and Gosnell 1924; Downs 1957; Riker and Ordeshook 1968;

Authors' note: Earlier versions of this article were presented at the Annual Meeting of the Western Political Science Association, Las Vegas, Nevada—The Riviera Hotel and Casino—March 8–10, 2007; and at the Annual Meeting of the Midwest Political Science Association, Chicago, Illinois—The Palmer House Hilton—April 3–6, 2008. We are grateful to Gregory Robinson, Jonathan S. Krasno, Christopher Zorn, Craig Laramee, Jasjeet Sekhon, Robi Ragan, Henry E. Brady, Cynthia Van Maanen, Superintendent Mark Capobianco, Dr. Annamary Allen, the Vestal Central School District, the Broome County Board of Elections, the Tioga County Board of Elections, Kevin Heard of the Binghamton University Geographic Information System Campus Core Facility, and Michael P. Welch and Blitwise Productions (www.blitwise.com). We are also grateful to the anonymous reviewers for their valuable feedback. We are grateful to the Harpur College of Arts and Sciences at Binghamton University and the National Science Foundation (Project: 1057309, Award: 40509). This research was approved by the Human Subjects Research Review Committee of Binghamton University on May 9, 2006; thanks to Anne Casella, vice chair of the committee, for her gracious support and assistance. All errors and omissions are the responsibility of the authors. Replication materials are available on the *Political Analysis* Web site.

Wolfinger and Rosenstone 1980; Rosenstone and Hansen 1993; Verba, Schlozman, and Brady 1995; Green and Gerber 2008). Recently, much of that attention has been focused on the circumstances of voting, especially on the number and location of polling places that are open (Gimpel and Schuknecht 2003; Brady and McNulty 2004; Dyck and Gimpel 2005; Gimpel, Dyck, and Shaw 2006; Stein and Vonnahme 2008). Despite evidence that ''accessibility'' affects turnout, eliminating polling places to cut costs has become an increasing trend across the United States.[1] The ''consolidation'' of polls simply means that a given administrative unit opens fewer polling places than it has in the past, aggregating multiple polling places into a smaller number of consolidated ones. Fortunately, many consolidations simulate a natural experiment in that they appear to affect would-be voters at random, a situation that offers special leverage to political scientists who study the impact of accessibility on voting. This article extends Brady and McNulty's earlier analysis of the effect of poll consolidation on turnout in Los Angeles County during the 2003 gubernatorial recall election by examining consolidation in a much different electoral setting, a school budget referendum in Vestal, New York.

The circumstances of the two elections could hardly be more different. The 2003 recall was conducted in the midst of a storm of publicity, almost uniformly critical, of the recently reelected incumbent, Governor Gray Davis (D-CA). One indication of the high interest in Davis' fate came on election day when 9,413,494 Californians (61.2% of registered voters) voted in a special election held on October 7, 2003, compared to 7,738,821 (50.6% of registered voters) who voted in the regularly scheduled election held just 11 months earlier. Turnout in Los Angeles County also rose from 44.9% of registered voters (in November of 2002) to 55.1% in the 2003 recall.[2] By contrast, school budget referenda, like most local elections in the United States, are held with less fanfare and lower turnout compared to their state and federal counterparts (Alford and Lee 1968; Morlan 1984; Hajnal and Lewis 2003). These relatively sleepy affairs (especially compared to the recall) are held every year in May, as prescribed by New York state law. Among voters registered for the 2005 election in the Vestal Central School District (VCSD), for example, there were 2340 ballots cast in the 2005 school budget referendum (38.9% of those registered) and 2171 ballots cast in the 2006 referendum (36.1%).

Los Angeles County and the VCSD do, however, have one important characteristic in common: both underwent poll consolidation. Los Angeles County reduced its number of polling places from 5231 to 1885 prior to the 2003 recall, and, on a much smaller scale, the VCSD reduced its eight polling places to five prior to the 2006 school budget referendum. Brady and McNulty (2004) report an approximately three percentage point decline in polling place voting in consolidated voting precincts compared to unconsolidated ones.[3] More specifically, Brady and McNulty find that the ''information cost'' accrued as a result of consolidation (whether one's polling place was changed) accounts for more of this decline than the ''transportation cost'' (the change in distance a person must travel to the poll from one election to the next).

[1]This is generally done as a means of saving money and streamlining administrative processes, an appealing option for perpetually underfunded boards of election mandated to make technological improvements in the wake of Florida's butterfly ballot problems in the 2000 presidential election.

[2]Source: California Secretary of State (http://www.sos.ca.gov/elections/elections_elections.htm).

[3]Brady and McNulty note that a substantial portion of their decline in polling place voting is recovered by a pick up in absentee voting. In the school budget referendum examined in this article, however, there are very few absentee ballots cast, even after the consolidation of polling places. As such, we focus on the decline in polling place voting throughout the article.

There is reason to believe a similar result may not be obtained in the VCSD. Most notably, the makeup of the electorate is quite different. In Los Angeles, turnout was driven by the controversy involving the sitting governor—the sort of high-profile issue to bring many new and occasional voters to the polls (Simon 2003)—and the outsize media attention and number of celebrity candidates (including the eventual winner) that were attracted to the virtually unprecedented[4] recall. In Vestal, however, it is fair to characterize the electorate as a particularly motivated one, where those who *do* vote do so rather habitually. For instance, of those who voted in the 2005 referendum (i.e., before the consolidation or "treatment" took place), more than 98% voted in the 2004 presidential election and approximately 90% voted in the 2002 midterm elections. Indeed, even the group of people that did *not* vote in the 2005 VCSD referendum, appear to be motivated voters: 95% of these registrants voted in the 2004 general election and around 80% voted in the 2002 general election. By any standard, this group of voters is about as reliable as one is likely to find in the United States.

This is important primarily because it could be the case that a highly motivated electorate is less likely to be affected by poll consolidation. Whatever it is that drives school budget referenda voters to the polls, be it civic duty or some other underlying motivation, the same factor might also make them especially likely to make the extra effort to find out where their new polling place is located and travel additional distance to the poll if that is required of them. Thus, among potential VCSD voters, polling place consolidation may not have (as much of) an impact on turnout as Brady and McNulty (2004) find in Los Angeles County. In contrast, if the consolidation does drive turnout down in this especially dedicated electorate, it will be a powerful demonstration of the effect of increases in information and transportation costs.

In the next section, we present our data and through randomization checks determine that a matching analysis is suitable for analyzing this particular natural experiment. We then present our results, which somewhat surprisingly show a more substantial decline in turnout as a result of consolidation than Brady and McNulty find. Finally, we conclude with a discussion of the implications of this finding.

1 The Electoral Setting and Data

Prior to 2006, there were eight polling places in the VCSD, shown in Fig. 1. First, there were polls set up in each of the district's five elementary schools. Then, for citizens who for the most part lived relatively far from any of the extant elementary schools, polls were opened in three conveniently nearby fire stations, as depicted by the gray circles and black triangles, respectively.[5] In 2006, the Vestal schools chose to close the polls located at fire stations, which

[4]The only other successful recall in the United States prior to California's in 2003 occurred in 1921 in North Dakota.

[5]The fire station polling sites were an artifact of previous districts that existed decades ago, when there were eight elementary schools in the VCSD. When the district decided to close three of the schools, there was significant resistance. As a concession, the school district agreed to maintain the old district lines from the eight elementary schools for voting purposes, with the polls at the defunct schools relocated to nearby fire stations. As the years passed, this created confusion among citizens who did not know why there were two conflicting district maps; it was finally decided in 2006 that enough time had passed since the concession to eliminate the old voting districts and adopt new ones contiguous to the elementary school attendance districts as they currently stand. This shifted voters not only away from the fire station voting districts (i.e., the districts where elementary schools were closed) but also shifted voters who were in neighborhoods that were in one still extant elementary school's district decades ago, but now were in another elementary school district as lines were redrawn over the years. In spite of the passage of so many years, there were still a few complaints (Source: Mark Capobianco, VCSD superintendent).

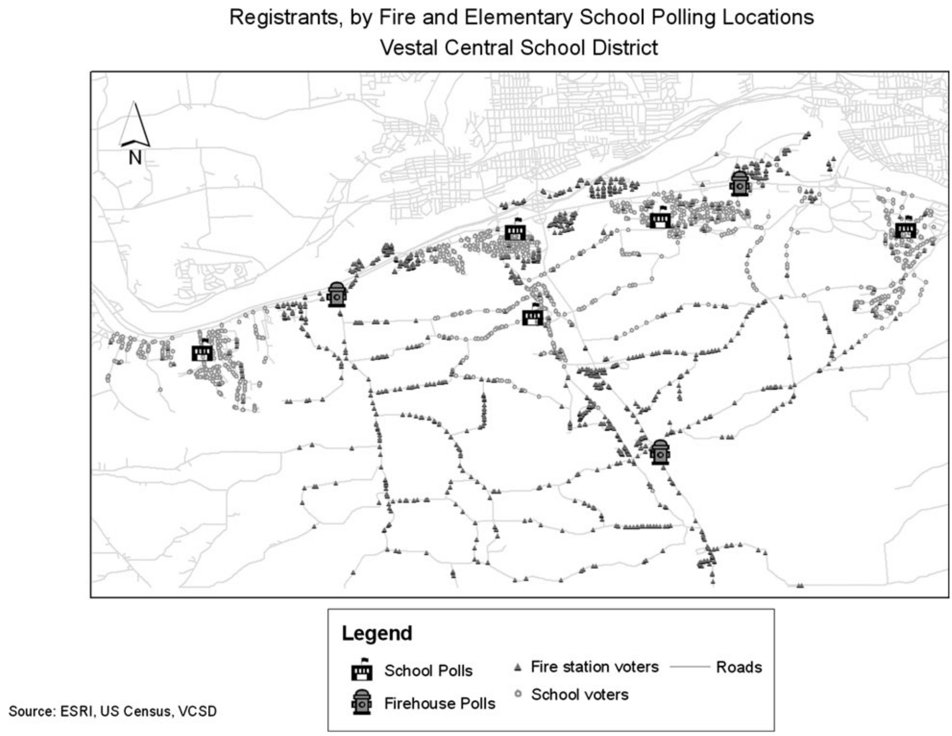

Fig. 1 The Vestal Central School District.

reduced the election budget by approximately $2580 and simplified election districting by making it contiguous with current elementary school attendance districts—all citizens are now instructed that their polling place is at the elementary school that serves their residence. This consolidation creates a natural experiment analogous to Brady and McNulty's (2004) in which residents of the VCSD who have their polling place changed incur an *information cost* and *transportation cost*.[6] If Brady and McNulty's results from Los Angeles are generalizable, we should observe lower turnout among those registrants whose polling place was changed, relative to those whose polling place was not changed. However, as outlined above, the particularly motivated nature of the VCSD electorate may result in disparate findings.

Our interest is in voters who were on the registration rolls in the VCSD of New York State in both May 2005 and May 2006. Our complete voter file contains 6012 individuals,[7] but of these, 213 were new registrants in 2006 and 195 dropped off between 2005 and 2006. Finally, using *ArcMap* geographic information system software, we were able to geocode (i.e., find the latitude and longitude coordinates of) the addresses of 5584 of 5603 (99.66%) of those individuals who were on the rolls in both 2005 and 2006. However, only 4566 (81.77%) of those individuals can be matched to a county file at their current address. So that we are able to control for other factors that are known to contribute to one's propensity to vote in the analyses that follow, these 4566 individuals are used to test whether

[6] For the interested reader, see Brady and McNulty (2004) for an extended discussion and formal model regarding the relationship between polling place consolidation and voter turnout.

[7] The rolls were recently restocked with new registrants, as there was a special election in March 2005 to approve funding for an expensive capital campaign to build or renovate several structures, most notably a new football stadium.

poll consolidation affects voter turnout.[8] Like Brady and McNulty (2004), our measure of the information cost is simply whether or not an individual's polling location was changed and we proxy the transportation cost by geocoding the original 2005 and the new 2006 polling place locations and calculating the change in distance[9] between the locations of voters and their polling places.

1.1 *Randomization Checks*

The decision to eliminate three polling locations for the 2006 VCSD election, unsurprisingly, resulted in an increase in the average distance that people had to travel to their polling locations. Figure 2 depicts this change in distance between 2005 and 2006 graphically. The histograms show how the consolidation added distance primarily to registrants on the high end of the distribution, pushing the shape of the normal curve down and to the right. Although we see commute distance increasing throughout the distribution, the bulk of the extra cost falls on registrants who already travel greater than average distances. It so happens that most of these registrants are former fire station voters (see Fig. 1); therefore, it appears that distance to the polls is one observable characteristic on which our natural "experiment" was not random.

More specifically, roughly one VCSD registrant of three (35%) was assigned to vote at a fire station in 2005, whereas the remainder went to their local elementary school (see Table 1). Those voters whose polling places were changed had to travel further to their polling place on average; those precincts were more sparsely populated and geographically remote than the school precincts. Thus, it is somewhat expected that the base turnout rate in 2005 for the voters whose polling places were changed is lower than the rate for those whose polling places were not changed; we can be 99.7% confident the observed difference is real and preexisting differences in the dependent variable forestall a true natural experiment (i.e., where the assignment mechanism creates statistically comparable groups by a random or quasi-random process.)

Therefore, preexisting differences must be controlled for, and covariates must be taken into account. Still, with some adjustments, similar groups can be compared and contrasted with regard to the exogenous intervention. As Robinson, McNulty, and Krasno (2009) argue in this volume, confining the use of natural experimental approaches to instances where the randomization is demonstrably perfect is far too restrictive a condition; arguably it is never appropriate under those conditions. Rather, the exogenous circumstance or incident, in this instance, polling place consolidation, that divides the population under study need be plausibly random, possibly with statistical correction.[10]

[8]Two points should be made in defense of this decision. First, the substantive findings of our analyses remain the same—the statistical significance or insignificance of independent variables on turnout do not vary depending on the presence of the 1018 school registrants who could not be matched to a county file (results available upon request). Secondly, much of these 1018 are "deadwood" on the school files, which are not routinely updated and purged of obsolete names as the county files are; the turnout rate in 2005 and in 2006 for this group are significantly lower (25.5% and 17.6%, respectively) than the turnout rates of those registered in the county. Only a small proportion of registrants register for school elections in advance at the administration building; the vast majority register at the polling place. It seems highly improbable that more than a handful of people actively and regularly participate in the school elections while failing to register for general election contests. It is probable that we failed to match some of these people properly for some reason, in what should be a random process. Hence, the inclusion of these school-only registrants, treating them as people who presumably chose to abstain, strikes us as risk averse to a fault and will increase random error in our observations and analysis (despite a larger N).

[9]Distance can be measured a number of different ways. See the Appendix for an extended discussion of the methodological and practical considerations involved.

[10]For another view, see Dunning (2008).

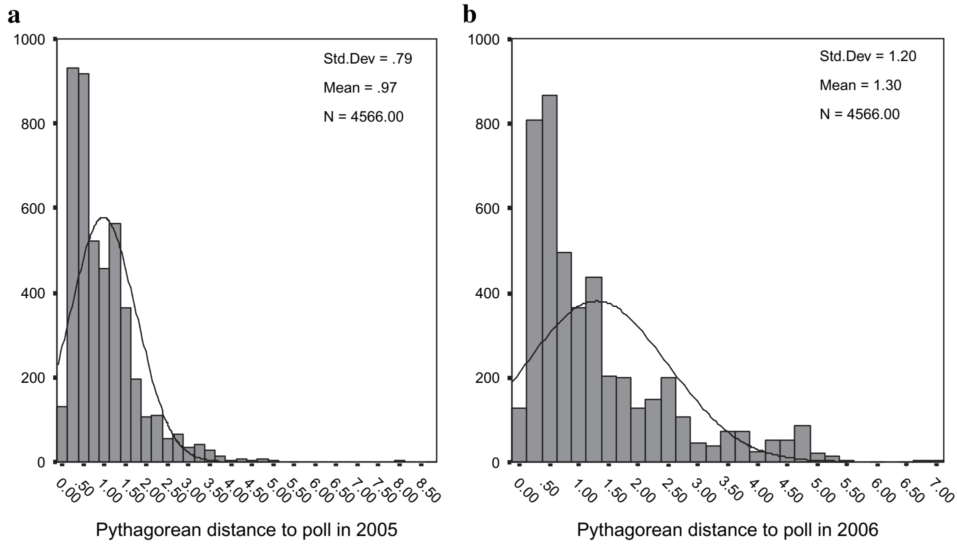

Fig. 2 Distance to the Polls, 2005 and 2006.

In this case, the reason for the selection effect is clear. The school district eliminated all (three) fire stations purposively, rather than randomly, to save money and streamline operations; given the experience of past consolidations, they were hopeful that any resulting decline in turnout would be small enough to be acceptable collateral damage. Indeed, the turnout rate at each individual fire station in 2005 was lower than that at any of the schools.[11] There are several reasons this may be the case. One likely contributing factor is that the mean distance traveled from one's home to the firehouses in 2005 was greater

Table 1 Results for 2006 and 2005 by polling place change, linked to county files

	Total electorate	*2006 turnout*	*2005 turnout*
Voters with changed polling place, *n* (%)[12]	1588 (35)	572 (32)	610 (32)
Turnout, %		36.0	38.4
Voters with unchanged polling place, *n* (%)	2978 (65)	1206 (68)	1280 (68)
Turnout, %		40.5	43.0
Total people, *n* (%)	4566 (100)	1778 (100)	1890 (100)
Turnout, %		38.9	41.4
2005 changed turnout minus 2005 unchanged turnout, %		−4.5**	−4.6**

*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$ (two-tailed test).

[11]Table 2 displays the specific turnouts for each polling place under both the 2005 and 2006 districting plans.
[12]Most of the people in this "changed" category voted at fire stations in 2005; however, a significant minority was switched from one school district to another.

than the mean distance to the schools, which were in more densely populated parts of the district. In addition, locating a poll at a school represents a more obvious focal point for a school district election than locating it at a firehouse.

One way of observing the preexisting differences between the precincts is by looking at the turnout rates by precinct in 2005 and 2006, and considering the opposite case both times—as we do in Table 2. The first difference between the elections, as one may have noticed in Table 1, is that the overall turnout rate is higher in 2005 than in 2006; the difference is statistically significant, with a confidence level exceeding 95%. Furthermore, the distinction between the fire station voters (the vast preponderance of the consolidated voters) and the school voters is robust. Looking at the 2005 contest, we see that turnout rates in the fire station precincts are consistently lower than in the school precincts, and if the consolidation had taken place in 2005, we would expect lower turnout at all the schools. The difference is somewhat slight, however, with the fire stations within a percentage point of the lower performing schools. In 2006, however, the gap widens considerably; those voters registered in "old" fire station precincts report to their new polling places in the schools notably less than those voters who remained in their original precincts. The worst performing school precincts have turnout rates in the mid-thirties, whereas none of the fire station precincts have turnout exceeding the low thirties.

We also estimated selection equations on our binary assignment variable. Specifically, we executed a logistic regression using the assignment variable (whether one had his or her polling place changed or not) as the dependent variable and testing all plausible independent variables in our data (see Heckman 1979; Achen 1986). There are statistically and substantively significant relationships with several critical independent variables (see Table 3). Most of the selection effect, however, is driven by one relationship: whether one was consolidated and the distance one had to travel in 2005 to one's polling place. The pseudo-R^2 of the selection equation regressing change in poll location on the 2005 distance alone is 25.0% (column [3]); with all other significant variables (age, minor party registration,[13] county voting history, and school voting history), the pseudo-R^2 is a miniscule 0.5% (column [1]).[14]

Therefore, we need to adjust for all of these variables, but especially for the distance variable, statistically. The variables identified in the selection equations (again, especially commute distance) are too highly correlated with the variable of assignment, whether one was consolidated or not, to permit us to confidently treat this analysis like a natural experiment. Assignment was not exogenous relative to other covariates, and hence, there is insufficient overlap between treatment and control on those covariates (past voting history, distance to poll, age, party identification, etc.). The voters, and the polls where the electorate vote, are not distributed evenly throughout the town; there are predictable neighborhood variations owing to the era of development and proximity to main roads.

---

[13]This is a peculiar, if trivial, result. Fewer than 3% of the school district registrants matched to county data are registered in minor parties (as opposed to Republicans, Democrats, and people with no party affiliations). This includes scattered memberships in parties of both the left (Greens, Working Families, Liberals) and the right (Conservatives, Libertarians, Independence Party members). These registrants were far more likely to be fire station voters than Democrats, Republicans, or Independents. It is hard to imagine this is anything other than coincidental; this group is disproportionately young, as is typical of minor party registrants, so our best guess is that this is a side effect of the uneven distribution of age brackets in Vestal.

[14]Age, in fact, proves insignificant without the presence of distance to poll. This is an artifact of heavy correlations with all the other variables in the selection model: a negative correlation with distance—the older more established parts of Vestal are near the schools; the newer construction is further away—and a heavy positive correlation with vote history in both the school district and the counties.

Table 2 Turnout by polling location, linked to county files

	Schools					*Fire stations*			
	Clayton Ave.	*Tioga Hills*	*Glenwood*	*African Road*	*Vestal Hills*	*Ross Corners*	*Vestal Center*	*Willow Point*	*Total*
2006, *n* (%)									
Observed	**417 (41.2)**	**306 (38.3)**	**414 (37.9)**	**305 (40.3)**	**336 (37.0)**	**—**	**—**	**—**	**1778 (38.9)**
Under 2005 precinct lines	*229 (41.6)*	*255 (39.4)*	*221 (44.2)*	*368 (41.4)*	*237 (38.3)*	*144 (33.9)*	*211 (34.1)*	*113 (32.2)*	*1778 (38.9)*
2005, *n* (%)									
Observed	**279 (50.6)**	**261 (40.3)**	**202 (40.4)**	**410 (46.2)**	**242 (39.1)**	**163 (38.4)**	**217 (35.1)**	**116 (36.6)**	**1890 (41.4)**
Under 2006 precinct lines	*499 (49.4)*	*318 (39.8)*	*407 (37.3)*	*332 (43.9)*	*334 (36.8)*	—	—	—	*1890 (41.4)*

Note. Values in bold indicate actual turnout as observed. Values in italics indicate turnout recalculated under the opposite districting plan.

Table 3 Selection regression results: Was polling place changed?

	(1) *Without 2005 distance to the poll*	(2) *With 2005 distance to the poll*	(3) *With only 2005 distance to the poll*
2005 distance to polling place		2.051 (0.069)***	2.043 (0.069)***
Age (in years)	0.003 (0.002)	0.005 (0.003)*	
Minor party registration	0.473 (0.176)***	0.408 (0.209)*	
VCSD turnout in 2005	−0.234 (0.065)***	−0.142 (0.077)*	
Vote history (county)	0.067 (0.023)***	0.089 (0.028)***	
Constant	−1.027 (0.135)***	−3.384 (0.181)***	−2.686 (0.078)***
Observations	4566	4566	4566
Pseudo-R^2	0.005	0.256	0.250
Prob > χ^2	0.000	0.000	0.000
Log likelihood	−2934.942	−2198.835	−2212.498

Note. Logit coefficients reported with SEs in parentheses.
*Significant at 10%, **significant at 5%, ***significant at 1%.

In the latter portion of this article, we adjust for these preexisting differences via a matching protocol.[15] We match on a number of observed covariates, so that within those covariates we simulate true randomization. It is therefore likely (although never certain) that unobserved covariates will be similar in the control and treatment group as well; this is the assumption of conditional ignorability of treatment. Putatively, the relationship between the treatment and turnout can be observed with a greater degree of confidence that any error in estimation is not due to biases caused by confounding or multicollinear variables. In this simulated experiment, we can see how the cost function of voters operates in a hermetic environment. First, however, we observe how poll consolidation affected turnout in the infinitely messier "real world."

2 Results

2.1 *Unmatched Analysis*

In order to estimate the impact of polling consolidation on voters in 2006, we model the decision to vote (scored 1) or not (scored 0) in 2006 by means of a logistic regression.[16] One independent variable—our measure of the information cost—of interest is whether the

[15]We follow the approach recommended in Imbens (2003). We have sufficient data that we can generate a data set of paired individuals that do not differ on any variable other than the assignment variable; in the one instance that it does, the difference would tend to bias against the expected relationship; the match balance is discussed in detail in Table 7. Arceneaux, Gerber, and Green (2006) demonstrate that under these circumstances, it is unnecessary (and indeed undesirable, given the arbitrary choices mandated in formulating the model) to employ propensity score modeling. In addition, Heckman, Ichimura, and Todd (1997, 1998), Heckman et al. (1998), and J. Smith and Todd (2001) suggest that data of the type that is typically generated by the voter rolls is especially amenable to this matching approach.

[16]In this case, there should be no distinction between whether we estimate the average treatment effect or the average effect of the treatment on the treated. The treatment here—residing in a zone where one's polling place was changed—happened to everyone in the treatment group, whether they were aware of it or not. Indeed, this awareness or lack thereof is a large part of the causal mechanism underlying the relationship between poll consolidation and turnout.

polling place was changed (scored 1) or not (scored 0) for the individual. Of the 4566 individuals in our data set, 1588 (34.8%) had their polling place changed. The other independent variable of the most interest to us—our measure of the transportation cost—is a measure of the change in distance to the polling place placed upon the individual.[17] This variable was constructed by first geocoding the addresses of both the polling locations (eight in 2005 and five in 2006) and those registrants[18] on the rolls both years, and calculating straight-line distances for both 2005 and 2006. Then, the 2005 distance was subtracted from the 2006 distance. This leaves us with a variable of changed distance where a positive number indicates that an individual's new polling place is further from his or her previous polling place, zero indicates there was no change in polling place,[19] and a negative number indicates that an individuals' new polling place is closer than his or her previous polling place.

There is obviously a strong relationship between these two variables empirically; if one's poll is not consolidated, then the value of both variables is null, and if one's poll is consolidated, the binary assignment variable has a value of one, and the distance variable is some nonzero number. One could observe them together using a "dosage model," with the mere existence of the nonzero distance figure indicating membership in the treatment group, and with the size of the distance change indicating magnitude. This would have the added advantage of being more econometrically efficient and reduces the risk of a biased estimate.

Theoretically, however, Brady and McNulty (2004) argue that the decision-making process and obstacles that must be overcome are too different to use a model with such a combined, continuous variable. Distance from the polls translates to transportation cost, which is effectively time lost *on* Election Day. Having a polling place change, however, incurs information cost—having to learn where the poll is, knowing who and what are on the ballot, and possessing at least a mild interest. The costs here are time lost prior to Election Day building up all these resources. A dosage design would commingle these effects and credit most of the variation to the effect with greater variance, the transportation effect. Therefore, although the concepts are related and may stem from a similar source (the consolidation), we model them separately given their theoretical underpinnings. We also present dosage models, which should confirm our primary models, with similar coefficients for the covariates and a weighted additive coefficient for the dosage variable.

The logit model also includes three controls for an individual's propensity to vote. Two of these relate to voting history. It is critical to control for past voting behavior because it is the chief indicator of the unobservable variable of political interest/

[17]To be perfectly correct methodologically, one would model these data hierarchically, with the poll change dummy variable on the second level, its variance determined by the number of precincts, and the remaining variables on the first level, with their variance determined by the number of registrants. However, the level 2 variable would only have eight cases, which is exceedingly small and leads to unreliable estimates. Nevertheless, we did model this in a hierarchical logit and we obtained results that are virtually identical to what we present in Tables 4 and 6. About 227 cases were dropped; the vast majority of these were from two specific neighborhoods, Castle Gardens and Twin Orchards, which were assigned to one elementary school under the old defunct eight-school plan and another under the new five-school plan. These neighborhoods have particular characteristics that make them noteworthy (they are older, less affluent communities built in flood zones); there is reason to believe excluding them could bias the results. The remaining cases were data updates or corrections.

[18]As mentioned above, a handful of registrants (22) could not be geocoded. These people either lived in newer developments that our address registries predated (17) or lived in trailer parks that lacked precise demarcations for the lots within their boundaries (5).

[19]It is also theoretically possible, albeit unlikely, that zero could indicate equidistance between the old and the new polling places.

Table 4 Logit results for turnout in 2006, linked to county files and controlling for vote history

	Coefficient	*SE*	*Z score*
VCSD turnout in 2005	1.602***	0.067	23.89
New polling place in 2006	−0.066	0.079	−0.84
Change in distance (miles)	−0.135***	0.039	−3.45
Vote history (county)	0.228***	0.024	9.30
GOP registration	0.160*	0.067	2.38
Constant	−2.323***	0.133	−17.48
No. of observations = 4566			
Prob > χ^2 = 0			
Pseudo-R^2 = 0.14			
Log likelihood = −2636.949			

Note. Logistic regression coefficients with SEs. The model was also estimated with fixed effects for each (2005) polling location. These fixed effects did not substantially improve the fit of the model and the statistical and substantive results of the independent variables were unaffected. For ease in interpretation, we have omitted the fixed effects from our report. In addition, clustering the SEs by 2005 polling location does not change any statistical or substantive conclusions. All of these results are available from the authors upon request.
*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$ (two-tailed test).

engagement. First, we observe whether or not the individual voted in the May 2005 school election and assign the registrant a score of one (1) if the individual did vote and zero (0) if the individual did not. Second, we generate a vote history variable based on general election participation (i.e., not school district elections). This variable is a count of the number of general elections the individual voted in since 2000 according to the county voter files; it ranges from zero (0) to six (6).[20] Finally, we also observe that Republican Party registrants vote in VCSD elections at a higher rate than registrants of the other parties (41.1% to 36.3% in 2006; 42.4% to 40.3% in 2005); hence, we controlled for Grand Old Party (GOP) registration.

Table 4 presents the results from the unmatched logistic regression of VCSD turnout in 2006.[21] The best predictor of future behavior is always past behavior, so as one would expect, 2005 turnout and the more comprehensive vote history variable are highly predictive of 2006 turnout. In addition, GOP registration increases the likelihood that an individual votes, relative to registration in any other party or registration in no party.[22] More importantly for our purposes, the variables relating to the consolidation both point in the expected negative direction, but only the change in distance to the polls is statistically significant.

Why does the transportation cost seem to have affected voters, whereas the information cost did not? We know that there is a strong relationship between one's distance from the poll and the assignment variable; one was much more likely to be consolidated if one was further from their poll initially. Therefore, those sustaining an increased information cost are also facing a debit on their already inflated transportation cost. This creates inefficiency

[20] We also controlled for age and major party identification without regard to party, but when the more comprehensive vote history variable is included, these controls do not attain statistical significance (results available upon request). The vote history measure captures the same variance as the age and party identification variables. We choose to present the more parsimonious model.
[21] Table 5 provides the summary statistics.
[22] This is a modest but statistically significant finding, which is not especially germane to our research agenda. We suppose this slight advantage in turnout among Republican registrants is an artifact of the heavily Republican lean of the VCSD, but we have no evidence supporting this supposition.

Table 5 Summary statistics for reported logit model, linked to county files and controlling for vote history

Variable	*Mean*	*SD*	*Minimum*	*Maximum*
VCSD turnout in 2006	0.3881	0.4874	0	1
VCSD turnout in 2005	0.4139	0.4926	0	1
New polling place in 2006	0.3478	0.4763	0	1
Change in distance (miles)	0.3240	0.9703	−7.35	4.94
Vote history (county)	4.8329	1.5238	0	6
GOP registration	0.5180	0.4997	0	1

in the model and potential bias. In Table 6, we look at different versions of this model, both as a robustness check and to ensure against badly biased estimators. We see that in a dosage model (the middle column), the result is, as predicted, virtually the same, with the effect of the assignment variable augmenting the effect of the distance variable and the other covariates remaining essentially constant. Furthermore, in the model in the left-hand column, we set aside magnitude effects and present a pure treatment effect. Similar effects obtain, with the assignment variable becoming much larger in absolute value and statistically significant. Therefore, based on the results from these other designs, it appears that the collinearity is not substantially biasing our estimation method.

2.2 *Matching Analysis*

As alluded to earlier, the assignment of who was to be consolidated was not done at random. Therefore, the above results are contingent on the particular constellation of dynamics in the VCSD, in this instance the fact that consolidated voters reside further from Vestal's population centers than unconsolidated voters do. Given that information costs and transportation costs were so correlated, it is difficult to discern the degree to which one or both of these are driving the negative turnout effect. Hence, we employ a matching analysis to create a hypothetical "as-if random" consolidation so we may disentangle the highly correlated variables.

Matching[23] is a method that takes observed units that received a treatment variable, in this case poll relocation, and "matches" them with units that did not. This process creates paired sets of data with opposite values on the assignment variable and like values on other observed key characteristics, creating groups for comparison that are likely to be statistically comparable.[24] The variables that we matched on are those that were found to have a statistically significant relationship with the variable we assign to be the experimental

[23]We matched the data using the statistical program R; specifically, we used the genetic matching (GenMatch) package, a component of the matching package developed and described in Sekhon (n.d.).

[24]Although one must always allow for unobserved variation, matching assumes minimally conditional ignorability of treatment assignment and stable unit treatment value. For discussions of the theory and method behind matching, see Neyman (1923), Rubin (1974), Rosenbaum and Rubin (1983), Holland (1986), Rosenbaum (2002), Imbens (2003), Imai and van Dyk (2004), Brady and McNulty (2004), Gerber and Green (2005), Arceneaux, Gerber, and Green (2006), and any of a number of articles with James Heckman as the primary author (see footnote 14).

Table 6 Robustness checks of results presented in Table 4

	Without distance	*Without new poll*	*Without GOP*
VCSD turnout in 2005	1.606*** (0.067)	1.603*** (0.067)	1.601*** (0.067)
New polling place in 2006	−0.188** (0.071)	—	−0.058 (0.079)
Change in distance (miles)	—	−0.150*** (0.035)	−.135*** (0.039)
Vote history (county)	0.224*** (0.024)	0.227*** (0.024)	0.231*** (0.024)
GOP registration	0.161* (0.067)	0.158* (0.067)	—
Constant	−2.309*** (0.133)	−2.337*** (0.132)	−2.258*** (0.130)
No. of observations	4566	4566	4566
Prob > χ^2	0	0	0
Pseudo-R^2	0.133	0.135	0.134
Log likelihood	−2642.876	−2637.298	−2639.795

Note. Logistic regression coefficients with SEs in parentheses.
$*p \leq .05$, $**p \leq .01$, $***p \leq .001$ (two-tailed test).

treatment, a change in poll location (see Table 3). These are initial (i.e., 2005 and prior) distance to the polling place, history of voting in the school district, voting history in general elections, age, and minor party affiliation, which we amend slightly to match on all party affiliations.[25]

The matching package we used uses an evolutionary algorithm to determine which of all the possible sets of paired cases yields the smallest overall variance between the multiple factors matched upon. This is the matching balance; the ratio between the central tendency and spread of a variable in the treatment group, and the central tendency and spread in the control group. After balancing, a matched data set was produced with 2732 cases.

With these matched data, whether one's poll was changed or not in 2006 is formally divorced from the distance one traveled to his or her assigned poll in 2005. This process disproportionately dropped people with short commutes, as the mean distance increased by more than 60%; the areas around the fire stations were less densely populated, and hence, there were fewer of them to match. The results, presented in Table 7 with accompanying summary statistics in Table 8, show that, *ceteris paribus*, the information cost is important, both statistically and substantively, in addition to the other factors we previously identified—vote history, party registration, and transportation cost. In other words, once we disentangle the highly related variables of distance and assignment, we observe a result akin to Brady and McNulty (2004) even in this very different electoral setting.

Specifically, the results in Table 7 indicate that for the average voter in the VCSD, the consolidation resulted in approximately a seven percentage drop in expected turnout.[26] We also observe that the relative relationship of the two disparate types of costs incurred by potential voters, information costs and transportation costs, obtains in this study as it did in Brady and McNulty (2004). To wit, although both costs have statistically significant

[25]Specifically, we seek to match Democrat to Democrat, Republican to Republican, no party to no party, and other to other. This is probably imperfect with respect to the minor parties, but they represent less than 3% of the matched data set, so the substantive effect should be negligible.

[26]To arrive at this figure, we held all variables at their means or modes, including "change in distance" = .5857 (see Table 8), and calculated the difference in the predicted probability of voting when "new polling place in 2006" equals zero (control) and one (treatment) (36.88 − 30.01 = 6.87). See Fig. 3 for a visual depiction of this decline.

Table 7 Logit results for turnout in 2006, linked to county files and controlling for vote history, estimated using data set genetically matched on poll change, balanced on vote history, age, party, and 2005 distance

	Coefficient	*SE*	*Z score*
VCSD turnout in 2005	1.435***	0.090	15.91
New polling place in 2006	−0.309**	0.100	−3.08
Change in distance (miles)	−0.110*	0.045	−2.46
Vote history (county)	0.247***	0.035	7.16
GOP registration	0.720***	0.089	8.10
Constant	−2.483***	0.195	−12.75
No. of observations = 2732			
Prob > χ^2 = 0			
Pseudo-R^2 = 0.12			
Log likelihood = −1590.904			

Note. Matching done in R using the GenMatch genetic matching package; the treatment variable was whether one had a new polling place or not; the data were matched on 2005 poll commute distance, participation in 2005 school election, an index on general election voting history, age, and party affiliation. Excellent balance was achieved with county vote history (control/treatment [Ctrl/Tr] mean ratio = 101.2%), age (99.4%), party affiliation (100.2%), and 2005 poll commute (99.4%). The balance with 2005 school turnout was poorer at 84.5%, but the Ctrl/Tr ratio was inverted from the one before matching, as well as slightly improved (117.9%); the bias in the matched data will vitiate against the observed relationship, a sterner test than a true match. The logit was estimated in *Stata*; further details are available from the authors upon request.
$^{*}p \leqslant .05$, $^{**}p \leqslant .01$, $^{***}p \leqslant .001$ (two-tailed test).

negative effects on turnout, the impact of the additional information cost incurred by the consolidation is much greater substantively than increases in transportation costs incurred by longer poll commutes; it would take an increase in distance of about 3.5 miles to achieve a drop in the likelihood of voting of about seven percentage points (see Fig. 3).

These results also highlight the criticality of using the matching method to correct for nonrandomness. The poll change and change in distance variables are inextricably related in the unadjusted data set; one cannot have a change in transportation cost without an accompanying change in information cost. The unadjusted analysis estimates the effect of both variables and assigns more weight to the variable that contains both the information and the transportation costs.

Matching on observed covariates, however, creates comparable groups where distance is uncorrelated with the changes in poll assignment. For the distance function, matching on

Table 8 Summary statistics for reported matched logit model, linked to county files and controlling for vote history

Variable	*Mean*	*SD*	*Minimum*	*Maximum*
VCSD turnout in 2006	0.3777	0.4849	0	1
VCSD turnout in 2005	0.3364	0.4726	0	1
New polling place in 2006 (matched)	0.5000	0.5001	0	1
Change in distance (miles)	0.5857	1.1424	−3.24	4.94
Vote history (county)	4.9557	1.4327	0	6
GOP registration	0.5608	0.4964	0	1

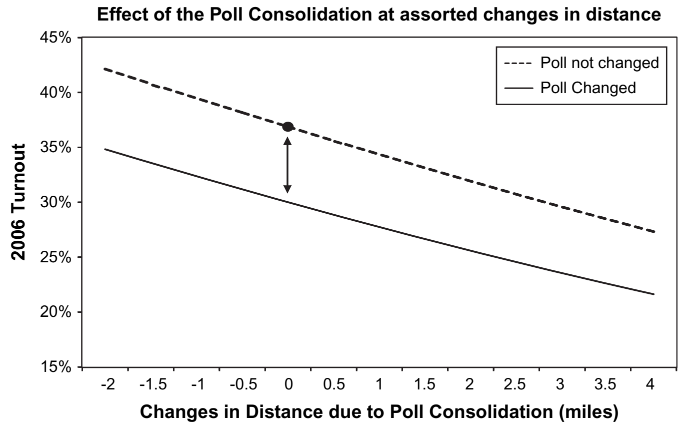

Fig. 3 Effect of the poll consolidation at assorted changes in distance.
Note. The dashed line, with the exception of the large black dot at $\Delta d = 0$, represents imputed unobserved estimations, as there are no observations in the data set with a nonzero change in distance when the polling location remains the same. The distance between the lines (highlighted with the up-down arrow) represents the information cost. The slope of the lines represents the transportation cost.

the distance of the 2005 commute lets us observe both the effect of the change and the effect of the magnitude of the change since 2005 poll distance is necessarily highly correlated with change in distance among the consolidated registrants. We also present a dosage design in Table 9 to demonstrate that modeling the assignment variable and the change in distance variable separately does not create bias in the estimate of the coefficients.

It can be assumed that there will always be some underlying turnout debit incurred when additional tasks or data collection (in this case, where to go vote) are added to the process of casting a ballot. This will be commingled or obviated by other factors in the real world, where myriad resources and costs are unevenly distributed throughout the electoral universe. Indeed, to the degree that an even distribution of resources and costs between treated and untreated populations is impossible to achieve in most electorates, and particularly in an electorate like the VCSD, which has low population density and spatial heterogeneity, the additional tasks may be a minor point. Nevertheless, these matched data indicate that the information cost remains the primary component of such a function in a *ceteris paribus* world, with transportation cost being a secondary component.[27]

3 Conclusions

Although the consolidation in the VCSD turned out to be far from a perfect natural experiment, using statistical matching methods we created substantial similarity (or "balance") between those who had their polling place moved and those who did not. The

[27]We also tested the matched data with a hierarchical model, noting the same caveats we outlined in footnote 17, to account for the nested nature of the data. These results, available upon request, bolster the conclusion that learning about the rules and structure for an election is more onerous than traveling to it.

Table 9 Logit results for turnout in 2006, using a Dosage model, linked to county files and controlling for vote history, estimated using data set genetically matched on poll change, balanced on vote history, age, party, and 2005 distance

	Coefficient	*SE*	*Z score*
VCSD turnout in 2005	1.409***	0.089	15.75
New polling place in 2006[28]			
Change in distance (miles)	−0.179***	0.039	−4.65
Vote history (county)	0.250***	0.034	7.27
GOP registration	0.719***	0.089	8.10
Constant	−2.599***	0.191	−13.59
No. of observations = 2732			
Prob > χ^2 = 0			
Pseudo-R^2 = 0.12			
Log likelihood = −1595.702			

*$p \leqslant .05$, **$p \leqslant .01$, ***$p \leqslant .001$ (two-tailed test).

matched data provide us with a very strong inference that consolidation in the VCSD reduced turnout by about seven percentage points in precincts in which the polling location was changed. We also found that, even in the different context, the change in polling place location had the same two effects as was found in Brady and McNulty's (2004) study of Los Angeles County: a transportation effect resulting from the change in distance to the polling place and an information effect resulting from the costs of finding and going to a new polling place. Moreover, as was the case in Los Angeles, it appears that the costs of information acquisition were more burdensome than the costs of travel to and from the polls but can be overcome with efforts by election administrators.

This finding brings to the fore another plausible interpretation of Brady and McNulty's (2004) results: that the three percentage point decline in turnout observed in Los Angeles is actually a lower bound given the recall's uniqueness, the celebrity (or notoriety) of some of the major candidates, including the eventual winner, and henceforth the media and public attention that inevitably surrounded it. Public attentiveness to the recall may have resulted in more information about poll consolidation being sought out and obtained than would happen under different, more typical electoral circumstances. From this perspective, perhaps it is surprising that Brady and McNulty find any difference in turnout between those in consolidated and unconsolidated precincts. If this is the case, then election administrators would be wise to take pains to inform voters of the consolidation and where their new polling location is, as those in the VCSD did through multiple channels, if they wish to keep voter turnout levels from falling. We cannot prove, but strongly suspect, that the substantial efforts invested by officials of the school district—most notably, they sent a postcard to every resident informing them of the change and the last district newsletter before the referendum included a reminder that the only polling place locations were located at the district's five elementary schools—took to inform the citizenry of the change in poll location prevented a steeper

[28]Note that with the absence of the assignment variable the above is a dosage design; as discussed in the text, the results look the same as Table 5 except that the Δd coefficient here is approximately the sum of the value (−0.110) from the Table 7 Δd coefficient plus the value of Table 7's assignment variable weighted by the mean change in distance (0.5857 miles).

turnout decline.[29] The apparent success of these efforts ought to be pleasing to conscientious election administrators.

Furthermore, the consolidation of polling places in the far more humdrum VCSD referendum demonstrates that even habitual voters are susceptible to the effects of poll consolidation. Perhaps this is because the habitual voter must unlearn longstanding practices and adapt, and that in voting, as in all of life, old habits are hard to break. Also demonstrated here is the necessity and utility of an experimental approach to studying events that occur in a state of nature, in this case using matching to create a simulated reality where the transportation and information costs were severable.

Funding

The National Science Foundation (Project: 1057309, Award: 40509); Harpur College of Arts and Sciences at Binghamton University.

Appendix: Methodological Note on Distance Calculations

In studies of this sort that look at travel distance from residence to poll, there are several different methodologies for measuring the distance between two points. The most commonly used is straight-line reckoning. This method is attractive because it relies on nothing more than high school geometry. One takes the longitude and latitude coordinates of the residence and the poll and applies the Pythagorean theorem, $a^2 + b^2 = c^2$, converting it to $c =$ the square root of $(a^2 + b^2)$, where a is the change in longitudinal coordinates between residence and poll, b the change in latitudinal coordinates between residence and poll, and c the hypotenuse of the triangle created by a and b, in other words the straight-line distance from the residence to the poll. According to Chamberlain (1996), for latitudes less than 50°, computation errors of no more than 66 feet (or, 1/80th of a mile) result.

However, it is of course a major concern that people, with only occasional exception, do not travel in straight lines to the polls.[30] Different attempts have been made to correct this.

The first way is obvious to city dwellers and is commonly called Manhattan city block estimation or grid estimation—indeed, in a place like Manhattan that is laid out (north of the Village) in a nearly perfect north-south, east-west grid, this is virtually a perfect estimator: one adds the absolute value of the difference in the longitudinal coordinates to the absolute value of the difference in the latitudinal coordinates:

---

[29]The administrators of the 2006 election, Superintendent Mark Capobianco and then District Clerk Annamary Allen, report that they did not anticipate the closing of the fire station polls would have a detrimental effect on turnout. However, they did anticipate an increase in complaints and inquiries that would be a resource drain. Their ameliorative efforts were designed to minimize those complaints and inquiries. It was a fortuitous bonus that they also may have minimized the turnout decline among the registrants in the consolidated election precincts.

[30]One obvious instance where one does travel something approximating a straight line (curb to curb) is if one lives on the same street as the polling place. We also observe people traveling in something approximating a straight line if they live exceedingly close to the poll and travel on foot, cutting across blocks via bike paths or park land and the like. However, these cases are inconsequential and for the most part theoretically uninteresting. There is some interest in people walking to the poll, but it is primarily in people walking nontrivial distances (i.e., more than a few hundred yards).

$|a_r - a_p| + |b_r - b_p|$, where a is the longitudinal coordinate (with the subscript indicating *r*esidence or *p*oll) and b the latitudinal coordinate. The method was used in Gimpel and Schuknecht (2003) and Dyck and Gimpel (2005). Manhattan city block estimation has the benefit of being even simpler than straight-line reckoning but also has the same problem—most places are not laid out in a perfect grid, so city block estimation is just another rough estimate of the true travel distance and a less parsimonious estimate at that.

Modern geographic information system (GIS) technology can use street maps converted into shapefiles to calculate actual driving distances curb to curb. These distances theoretically represent the most precise estimation of distance. If one assumes a voter will travel by car and will go directly to the poll by the shortest possible route (both flawed but necessary assumptions), the only variable involves availability of parking spaces, by law a factor that must be minimized at the polling place, and which is in any event a matter of feet.

One calculates distances using GIS software by using the shortest network path extension in *ArcView* (software made by Environmental System Research Institute). This tool allows users to find the shortest route between two separate shapefiles that contain points, using another shapefile that provides the base network for the view. The network in this case is the lattice of streets on the shapefile map that shows possible paths between two points. The software calculates the shortest path between two points in the network and designates it the "cost" of moving between the two points. This cost is then appended to the origin attribute table in units that are determined by the reference units of the view in its entirety—-degrees of distance. These can be converted to feet, kilometers, or miles with little difficulty[31]—in any case the particular unit of measure is unimportant for comparative purposes.

The GIS software has another advantage over the derived estimates using Manhattan block distance and Pythagorean interpolation—it assumes a round earth. One can also figure straight-line distances taking the earth's curvature into account. Brady and McNulty (2004) calculated distances according to the most common estimator, the great circle method, which uses the law of cosines.[32] It can be summarized as follows:

$$d = 1.15077945 \times 60 \times \mathrm{ARCOS}\left[\mathrm{SIN}(b_r) \times \mathrm{SIN}(b_p) + \mathrm{COS}(b_r) \times \mathrm{COS}(b_p) \times \mathrm{COS}(a_{r-p})\right]$$

where,

d is computed distance,
b_r the latitude at the first point,
b_p the latitude at the second point,
a_r the longitude at the first point,
a_p the longitude at the second point,
a_{r-p} the longitude of the second point minus longitude of the first point,

[31] Here is one conversion protocol: 1° = (radians/PI) × 180; 1 min of arc is 1 nautical mile (and there are 60 min in a degree); from there, 1 nautical mile = 1.852 km and 1 km = 0.621371192 miles, or 1 nautical mile is 1.15077945 miles. In contrast, for distances over relatively small areas it should be sufficient to figure out the width of a degree of longitude at the latitude under observation, choosing arbitrarily either the start point or the end point (in this data set the largest observed effect of choosing one or the other varies the estimation by about 320 feet, about the length of a football field). One calculates the width of 1° of longitude in miles at a given latitude by taking the cosine of the latitude and multiplying it by 69.17, the width of a degree of latitude at the equator. Latitudinal degrees are an invariant 69.17 miles wide, but longitudinal degrees get smaller as latitudes increase because the distances to circumnavigate the globe decrease as one approaches the poles.

[32] http://www.ga.gov.au/geodesy/datums/distance.jsp

assuming,
1 min of arc is 1 nautical mile,
1 nautical mile is approximately 1.15077945 miles.

However, relying on the spherical law of cosines—in practice, the arccosine—results in large rounding errors for short distances. Simply put, the earth is so close to flat over short distances that the cosine function becomes asymptotically close to 1, so close that it overwhelms standard computer calculators' ability to accurately estimate the arc. Chamberlain (1996), citing Sinnott (1984), notes that "a computer carrying seven significant figures cannot distinguish the cosines of any distances smaller than about one minute of arc." Our informal observations imply that the problem extends to even greater distances, but accepting Sinnott's estimate, given that 1 min of arc is approximately 1.15 miles, this issue poses a problem for data sets in which a large number of the cases are (in reality) distances less than 1.15 miles. This would include virtually any data set looking at the U.S. elections of which we can conceive. Thus, great circle distance, calculated relying on the spherical law of cosines, is problematic because it cannot reliably distinguish among these shorter distances.

Another calculation method that takes the earth's curvature into account is the Haversine method, which uses the arcsine rather than the arccosine and is thus more accurate for shorter distances:

$$d = 1.15077945 \times 60 \times 2 \times \text{ARCSIN} \times \text{SQRT}\{\text{SIN2}[(b_r - b_p)/2] + \text{COS}(b_r) \times \text{COS}(b_p) \times \text{SIN2}(a_{r-p}/2)\}$$

where
d is the computed distance,
b_r the latitude at the first point,
b_p the latitude at the second point,
a_r the longitude at the first point,
a_p the longitude at the second point,
a_{r-p} the longitude of the second point minus longitude of the first point,

assuming,
1 min of arc is 1 nautical mile,
1 nautical mile is approximately 1.15077945 miles.

The Haversine estimation improves on the great circle estimation, reducing erroneous measures. One thing we know must be true is that a curved distance must be longer than a straight distance; if an estimate says otherwise, it is wrong. The great circle estimation, because of the imprecision in the cosines, has an error rate 15.8 times greater than the Haversine estimation. The Haversine estimation only has 32 cases like this in the entire set of 5779 used for this analysis (some cases were dropped because of missing data), and the biggest error is less than 15.5 feet, almost certainly attributable to rounding at some stage.

However, Table A1 shows that the choice of measurement method ultimately makes very little difference in the final analysis. What we observe first is near-perfect correlation between the geometric and the trigonometric estimates. This is unsurprising—any meaningful variation would necessarily be the result of some gross computation error. More relevantly, this demonstrates that the relationship of distance to turnout (or any other variables) should be virtually identical regardless of which measure is used.

Table A1 Correlation of different measurement methods

	Pyth	*Grid*	*GC*	*Hav-sin*	*Roads*
Pythagorean hypotenuse (flat earth)	1				
Manhattan block grid (flat earth)	0.986**	1			
Great circle (GC) estimation, law of cosines (round earth)	0.986**	0.986**	1		
Haversine estimation, law of sines (round earth)	0.993**	0.987**	0.998**	1	
Road map, street distance (Round Earth)	0.862**	0.859**	0.873**	0.874**	1

**Correlation is significant at the .01 level (two tailed).

The association of street distance to the other measures is somewhat weaker, ranging from 0.859 to 0.874, but is statistically significant with a confidence level exceeding 99%. This weaker correlation makes intuitive sense: roads and highways meander according to geographic contours, property lines, and various impedances, sometimes necessarily following very indirect routes from a given point A to point B. Indeed, that intuition motivated this analysis. However, the statistical significance of the association means that the extra precision of the street map distances should not yield appreciable differences in estimating the relationship of distance to turnout. Given this, a simple estimation like Pythagorean straight-line distance or Manhattan block distance is wise for researchers to adopt. They provide parsimony in computation and explanation, freeing the social scientist to focus on areas more germane to the theory at hand, without sacrificing analytic power or accuracy.

References

Achen, Christopher. 1986. *The statistical analysis of quasi-experiments*. Berkeley: University of California Press.

Alford, Robert R., and Eugene C. Lee. 1968. Voting turnout in American cities. *American Political Science Review* 62:796–813.

Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. Comparing experimental and matching methods using a large-scale voter mobilization experiment. *Political Analysis* 14:37–62.

Brady, Henry E., and John E. McNulty. 2004. The costs of voting: Evidence from a natural experiment. Presented at the Annual Meeting of the Society for Political Methodology, Palo Alto, CA.

Chamberlain, Robert G. 1996. Great circle distance between two points. October. http://www.movable-type.co.uk/scripts/gis-faq-5.1.html.

Downs, Anthony. 1957. *An economic theory of democracy.* New York: Harper & Row.

Dunning, Thad. 2008. Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly* 61:282–93.

Dyck, Joshua J., and James G. Gimpel. 2005. Distance, turnout, and the convenience of voting. *Social Science Quarterly* 86:531–48.

Gerber, Alan S., and Donald P. Green. 2005. Correction to Gerber and Green (2000), Replication of disputed findings, and reply to Imai. *American Political Science Review* 99(2):301–13.

Gimpel, James G., Joshua J. Dyck, and Daron R. Shaw. 2006. Location, knowledge, and time pressures in the spatial structure of convenience voting. *Electoral Studies* 25:35–58.

Gimpel, James G., and Jason E. Schuknecht. 2003. Political participation and the accessibility of the ballot box. *Political Geography* 22:471–88.

Green, Donald P., and Alan S. Gerber. 2008. *Get out the vote: How to increase voter turnout,* 2nd ed. Washington, DC: Brookings Institution Press.

Hajnal, Zoltan L., and Paul G. Lewis. 2003. Municipal institutions and voter turnout in local elections. *Urban Affairs Review* 38:645–68.

Heckman, James J. 1979. Sample selection bias as a specification error. *Econometrica* 47:147–61.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. Characterizing selection bias using experimental data. *Econometrica* 66:1017–98.

Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. Matching as an economic evaluator estimator: Evidence from evaluating a job training program. *Review of Economic Studies* 64:605–54.

———. 1998. Matching as an economic evaluation estimator. *Review of Economic Studies* 65:261–94.

Holland, Paul. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81:945–60.

Imai, Kosuke, and David A. van Dyk. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association* 99:854–66.

Imbens, Guido. 2003. Semiparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* 86(1):4–29.

Merriam, Charles E., and Howard Gosnell. 1924. *Non-voting*. Chicago, IL: University of Chicago Press.

Morlan, Robert L. 1984. Municipal versus national election voter turnout: Europe and the United States. *Political Science Quarterly* 99:457–70.

Neyman, Jerzey. 1923. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Edited and translated by D. M. Dabrowska and T. P. Speed. *Statistical Science* 5:465–72.

Riker, William H., and Peter C. Ordeshook. 1968. A theory of the calculus of voting. *American Political Science Review* 62:25–42.

Robinson, Gregory, John E. McNulty, and Jonathan S. Krasno. 2009. Observing the counterfactual: The search for political experiments in nature. *Political Analysis.* Advance Access published on August 5, 2009. 10.1093/pan/mpp011.

Rosenbaum, Paul. 2002. *Observational studies*. 2nd ed. New York: Springer-Verlag.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.

Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, participation, and democracy in America*. New York: MacMillian.

Rubin, Donald B. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66(5):688–701.

Sekhon, Jasjeet S. n.d. Multivariate and propensity score matching software with automated balance optimization: The matching package for R. Journal of Statistical Software. Forthcoming.

Simon, Mark. 2003. Recall spurs new voters to voter registration: Informal tallying shows numbers up despite summer usually being slow. *San Francisco Chronicle* August 21, Section A-1.

Sinnott, Roger W. 1984. “Virtues of the Haversine.” *Sky and Telescope* 68:158.

Smith, Jeffrey, and Petra Todd. 2001. Does matching overcome LaLonde’s critique of nonexperimental methods? *Journal of Econometrics* 125:305–33.

Stein, Robert M., and Greg Vonnahme. 2008. Engaging the unengaged voter: Voter centers and voter turnout. *Journal of Politics* 2:487–97.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and equality: Civic voluntarism in American politics*. Cambridge, MA: Harvard University Press.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who votes?* New Haven, CT: Yale University Press.