THE COSTS OF VOTING: EVIDENCE FROM A NATURAL EXPERIMENT

Henry E. Brady
hbrady@csm.berkeley.edu
Department of Political Science
University of California, Berkeley

John E. McNulty
mcnulty@socrates.berkeley.edu
Department of Political Science
University of California, Berkeley

ABSTRACT: *This paper uses the consolidation of polling places in Los Angeles County during the October 2003 gubernatorial recall election to study the costs of voting. The consolidation afforded an opportunity to observe a natural experiment: those whose polling places were changed in the consolidation should be statistically comparable to those whose polling places are not changed. Thus, we may observe both groups' turnout rates and ascribe any differences to the treatment of the change in polling locus. We find evidence that changing polling place locations does decrease turnout overall by a substantial 1.88 percentage points; a drop in polling place turnout of 3.05 percentage points is offset by an increase in absentee voting of 1.19 percentage points.*

Almost 50 years ago, Anthony Downs introduced the critical insight that the act of voting has costs, and that when the costs get sufficiently high, it is rational for a voter to abstain (Downs 1957). In theory, the cost need only be barely non-zero to justify abstention, so low is the probability that one's single vote will affect the outcome of the election and thus produce a benefit. In practice, of course, people do vote, demonstrating that the act of participation is not only an exercise in self interest, but also involves aspects of altruism, civic-mindedness, and expressiveness, thus justifying the personal cost in time, transportation, and inconvenience incurred (Riker & Ordeshook 1973; Green & Shapiro 1994).

However, these higher-minded motivations are not universally sufficient to overcome the costs of voting for all people; while turnout rates are very variable, they never approach 100% in any election of consequence. So, costs do matter to voter turnout. The challenge then is to identify these costs and quantify them, in order to build the best model we can of those factors affecting voter turnout and participation more generally.

The biggest determinants of participation involve traits of individual voters which affect costs or perceived costs of voting, such as political interest or education (Campbell, Converse, Miller, & Stokes 1960; Miller & Shanks 1995). However, accessibility issues and convenience factors have been shown to have significant effects (Rosenstone & Wolfinger 1978, Wolfinger & Rosenstone 1980; Squire, Wolfinger, & Glass 1987; Rosenstone & Hansen 1993; Knack 1995; Highton 1997; Stein 1998).

In recent work, Gimpel & Schuknecht (2003) investigated the specific question of whether the difficulty of reaching one's polling place, in terms of distance and impedance

(i.e., the time and effort of the commute). They determined in a 2000 general election study of three suburban Maryland counties that ease of access is positively related to turnout, though the relationship is non-linear and moderated by other factors.

The historic California gubernatorial recall election of 2003 provides an opportunity to study voter turnout and the costs of voting in what amounts to a unique natural experiment. The almost surreal event, governed by a dusty, century-old constitutional clause and covered by a blanket of international media, created major problems for county election officials who had not budgeted for the unexpected recall election. In seeking to cut costs, some counties consolidated voting precincts.

Consolidation was possible because of the unique ballot for the Recall election. In a typical election there are scores of different ballots in a county, based on overlapping lines of different types of governmental units for which elections are being held. This makes it hard to consolidate precincts without risking having multiple ballots in the same precinct – a confusing practice at best. In the special election held for the Recall, however, there were only four, statewide, questions on the ballot, the two part recall question and two initiatives. The first recall question was whether the sitting governor should be recalled, and the second was who, from a list of candidates, should replace him. The only ballot difference across areas was the need to randomize the order of candidates in each of eighty Assembly districts. Consequently, precinct consolidation was both feasible and a reasonable response to budgetary strictures.

Not every county consolidated precincts. In fact, most did not. Despite the cost factor, county administrators were loath to risk the possibility of a decline in voter turnout – and an increase in voter complaints – bound to occur given changes in long

established polling places and a decrease in the density of the polling places offered. There are two conspicuous ways that voter turnout could be decreased by a decrease in the polling places. First, new polling places could be less accessible to some voters because they are farther away, a "transportation effect". Second, new polling places require that the voter know that the polling place has changed, know how to get to the new location, and be undeterred by lack of knowledge about a new neighborhood. Problems related to these issues can be called a "disruption effect". County election administrations have as one of their primary goals the maximization of voter turnout. Since consolidation could at least theoretically lead to a less than maximal turnout, most shied away from it.

A few counties, however, consolidated precincts. Los Angeles County, the biggest county in California and in the United States (and thus a place that incurs some of the heaviest costs in absolute terms in election administration), reduced the number of distinct voting precincts from 5,231 to 1,885. This process in Los Angeles County created what we hoped would be a natural experiment with a large number of observations. We hoped that whether or not someone's polling place was the same or different between 2002 and 2003 would be completely uncorrelated with those characteristics that affect one's predisposition to vote. If this were true, then we could compare two groups, alike in every way, except that one group, the "treatment group," has had its "cost of voting" increased by the change in polling place, forcing the registrant to digest and cope with new information and challenges that the other group need not. The other group, the "control group," did not have to deal with these problems so that we

can estimate the impact of precinct consolidation from taking the difference in turnout between the treatment and control groups.

Although the consolidation is not a perfect natural experiment, it is about as close as we normally come with observational data. Consequently, it provides us with a very strong inference that consolidation in Los Angeles County reduced turnout by a substantial 1.88% in the precincts in which the polling location was changed. We also find that voting at the polling place decreases even more, by 3.05% but that an increase of absentee voting of 1.19% makes up for some of this reduction. In addition, we find that the substitution of absentee voting for a reduction in polling place voting is greatest among people of middle age and older whereas younger people are more inclined to simply not vote at all.

We also find that the change in polling place location has two effects: a transportation effect resulting from the change in distance to the polling place and a disruption effect resulting from the information required to find the new polling place and the risk aversion that people feel about going into a new neighborhood. The disruption effect is about five times larger than the transportation effect for the average person who experienced an increased distance to the polling place of about a sixth of a mile, but the effects were roughly equal for someone who had an increased distance of about a mile.

Theory

What Precinct Consolidation Does -- The consolidation of voting precincts consists of three distinct changes. First, for most people whose polling places are changed, the new polling place will be farther away than the old polling place. Hence,

there will be additional *transportation costs* for most people. There are also two kinds of *disruption* effects. One is the increased *informational and search costs* associated with finding the new polling place. The other is that the new polling place will typically be in a different neighborhood than the old one, and this neighborhood will usually differ from it in many respects. On average, it seems just as likely that the new neighborhood will be better (more safe, easier to get around, more pleasant to be in, etc.) rather than worse than the old one, but uncertain and risk-averse voters will weigh the costs more heavily than the benefits. Hence, on average, there will be increased *risk aversion costs*.

Our research program involves trying to disentangle these costs by getting information on each of them. Informational and search costs can be proxied by whether or not the polling place location was changed. Transportation costs can be proxied by geo-coding the original 2002 and the new 2003 polling place locations and calculating the increased distance from voters. Finally, risk aversion costs can be proxied by getting census data (or crime data) on the characteristics of the neighborhood of the polling places. The first measure, whether or not the polling place was changed, is relatively easy to get, and we have it for this paper. The second measure, distances to the old and new polling places, has been calculated and incorporated in this paper. We are still working on collecting neighborhood characteristics.

For some of the analysis in this paper, we just focus on whether or not polling place location was changed. When used alone, this variable will capture the average effects of all the processes described above, and it will allow us to answer the major question confronting those election officials who want to consolidate districts which is whether or not there will be any significant decline in turnout. If there is a significant

decline, then one policy response is for election officials to forego precinct consolidation. Another would be to encourage absentee ballots. But it would be useful to know what combination of additional informational and search costs, transportation costs, and risk-aversion costs caused the decline because each of these factors has different policy implications. If the problem is informational, then voters might be better informed about how to get to their new polling location. If the problem is transportation costs, then precinct consolidation might be undertaken to minimize the increase in the distance people must travel, some programs might be instituted to help people get to the polls (especially if the transportation costs seem to fall especially heavily on some portions of the population), and absentee voting might be encouraged. Finally, if the problem is risk aversion costs, then care might be taken to choose new polling places in good neighborhoods or if this is infeasible (due to distance considerations, for example), in "safe haven" sites in bad ones; also, voters might be given better information about the neighborhood in which the new polling place is located. Absentee voting might also be encouraged.

The Outcomes of Precinct Consolidation -- Turnout, the fraction of people who vote either at the polling place or by absentee ballot, is obviously an important outcome variable for this study, but it is not the only possible outcome measure. There are three possible outcomes for each voter, voting at polling places, absentee voting, and not voting, and turnout is the sum of the first two of them. To clarify what we expect from precinct consolidation, we consider each outcome represented by a letter with a value of one if the voter chooses that outcome and zero otherwise. Thus, the voter can vote at the *polling place* on election-day (represented by *p*), the voter can vote via *absentee* ballot

(*a*), or the voter can decide *not to vote* at all (*n*). Probably the most important policy

question is the impact of consolidation on voting turnout (*t*) which consists of voting at

the polling place or via absentee ballot (*t* = *p* + *a*).   Obviously the sum of either turning

out (*t*) or not turning out (*n*) must be unity because one or the other act must occur.

Hence, *t+n = 1*. With this identity and the definition of turnout, any two of the measures

p, a, and *n* provide a full description of a voter's behavior.

Our model considers how the costs and benefits of voting are related to polling

place, absentee, and non-voting. The costs of voting at the polling place are represented

by c^p (these costs consist of the sum of c^i—information and search costs, c^t –

transportation costs, and c^r –risk aversion costs) and the costs of voting absentee by c^a.

The benefits of voting are represented by *b*. Then the utility U_i^p that person *i* gets from

polling place voting *p* will be the net benefits b_i-c_i^p for voting at the polling place. The

utility U_i^a for voting absentee will be the net benefits b-c^a for voting absentee, and the

utility U_i^n for not voting will be 0.[1] Clearly, the person will maximize his or her welfare

by voting at the polling place (p_i = *1*) if the net benefits of voting at the polls is greater

than zero (b_i-c_i^p > *0*) <u>and</u> the net benefits of voting at the polls exceed the net benefits of

absentee voting (b_i-c_i^p > b_i-c_i^a). The person will vote absentee if the net benefits of

absentee voting are positive (b_i-c_i^a > *0*) <u>and</u> the net benefits of absentee voting are greater

than the net benefits of voting at the polls, (b_i-c_i^a > b_i-c_i^p). And finally the person will not

vote if the (zero) net benefits of not voting are greater than the net benefits of voting at

the polls (0 > b_i - c_i^p) and the net benefits of voting absentee (0 > b_i - c_i^a)

---

[1] Thus the voter's decision problem is to maximize the expression [(1-N)*B – (1-N)*A* C_A – (1-N)*P* C_P] by choosing P, A, or N.

If we think of b^i, c_i^p and c_i^a as realizations of random variables b, c^p, and c^a with a trivariate probability distribution that describes the voting population, then the proportion of people voting at the polls, for example, is equal to the following, where we have used capital letters to represent proportions:

(1) $P = \text{Prob}(p=1) = \text{Prob}(U_i^p > U_i^a \text{ and } U_i^p > U_i^n)$

$$= \text{Prob}(b-c^p > b-c^a \text{ and } b-c^p > 0) = \text{Prob}(c^a > c^p \text{ and } b > c^p).$$

And similarly for the proportion of people voting absentee:

(2) $A = \text{Prob}(a=1) = \text{Prob}(U_i^a > U_i^p \text{ and } U_i^a > U_i^n)$

$$= \text{Prob}(b-c^a > b-c^p \text{ and } b-c^a > 0) = \text{Prob}(c^p > c^a \text{ and } b > c^a).$$

And for the proportion of people not voting:

(3) $N = \text{Prob}(n=1) = \text{Prob}(U_i^n > U_i^p \text{ and } U_i^n > U_i^a)$

$$= \text{Prob}(0 > b - c^p \text{ and } 0 > b - c^a) = \text{Prob}(c^p > b \text{ and } c^a > b).$$

Figure 1 presents a graphical depiction of these formulas. Suppose we plot net benefits from polling place voting ($b-c^p$) versus net benefits from absentee voting ($b-c^a$). Each voter will be located somewhere in this space depending upon his or her values of b, c^p, and c^a. If a third dimension were added to this picture, it could represent the density of each kind of voter. The diagonal on the figure represents the place where net benefits from voting place voting equals the net benefits from absentee voting. The zero line on each axis represents the place where net benefits are zero. The non-voters, indicated by *N* on the picture, are in the lower left-hand quadrant where net benefits are negative for both polling place and absentee voting (where $0 > b - c^p$ and $0 > b - c^a$). The polling place voters, *P*, are above the diagonal ($b-c^p > b-c^a$) where the net benefits of polling place voting exceeds the net benefits of absentee voting, and they are above the zero net

benefits line for polling place voting ($b-c^p > 0$). The absentee voters are below the diagonal ($b-c^a > b-c^p$) and to the right of zero net benefits for absentee voting ($b-c^p > 0$). The proportion in each of the three groups depends upon the density of voters in each area.

Now, consider what happens with precinct consolidation through grouping. Assume that all precincts are grouped and the cost of grouping, c^g, is the same across all precincts. For voters the cost of voting at the polling place increases from c^p to $c^p + c^g$. Then the proportions change to the following:

$$P^* = \text{Prob}(c^a > c^p + c^g \text{ and } b > c^p + c^g).$$

(4) $\quad A^* = \text{Prob}(c^p + c^g > c^a \text{ and } b > c^a).$

$$N^* = \text{Prob}(c^p + c^g > b \text{ and } c^a > b).$$

The first formula indicates that polling place voting will unequivocally decrease if c^g is greater than zero and if there are people for whom this change makes a difference because there will be fewer people for whom the benefits of voting exceed the costs of polling place voting and for whom the costs of absentee voting are greater than the costs of polling place voting. That is, some people will move from polling place voting into absentee voting and others will move from polling place voting into not voting at all. Those who move into absentee voting will be people who always thought that the benefits of voting outweighed the costs of absentee voting but who formerly found it cheaper to vote at the polling place than through absentee ballots and who now find it better to vote absentee because of the added cost, c^g, to polling place voting. Those who move into non-voting will be those who never voted absentee (and won't now) because they calculate the costs of absentee voting to be greater than the benefits of voting, but

they voted in the past because they found the benefits of voting at the polls to be greater than the costs of voting there. With the additional costs of voting at the polls, and with their long-standing belief that absentee voting costs more than the benefits of voting, they decide not to vote at all.

Figure 2 represents these changes by making shifts in two lines. First, the diagonal line in Figure 1 shifts upward by the amount c^g because the net benefits of polling place voting have decreased by that amount. Consequently, the net benefits of polling place voting will now only equal the net benefits of absentee voting for those people for whom their net benefits of polling place voting used to be c^g units bigger than the net benefits of absentee voting. Second, the horizontal "zero" line shifts upward by c^g because the net benefits of polling place voting are now greater than zero only for those people for whom the benefits used to be c^g units bigger than zero. The resulting picture has two areas where voters move away from polling place voting. $N^\#$ are people who no longer vote. $A^\#$ are people who turn to absentee voting. The relative size of each group will depend upon the size of c^g and the density of voters in these areas. Note that using Figures 1 and 2 we can write the proportions of each kind of voter (P^*, A^*, and N^*) in the final situation in terms of the proportions in the original situation and those who change their behavior:

$$P^* = P - N^\# - A^\#$$

(5) $$A^* = A + A^\#$$

$$N^* = N + N^\#$$

We can also write turnout as:

(6) $$T^* = P^* + A^* = P + A - N^\#$$

This formula shows that the change in turnout will be the negative of the change in non-voting, and turnout will decrease less than polling place voting because some people will move away from polling place voting into absentee voting. Our goal is to quantify these effects and to relate them to the strength of the treatment as measured by the change in distance to the polling place and to individual characteristics.

Empirical Results

The Data – Conceptually, getting the data ready for this paper was straightforward, but the sheer size of the data files and lists involved made it anything but simple. We obtained voter lists, along with their addresses and precincts, for both the 2002 (Gubernatorial and midterm election) and the 2003 recall election. We also obtained lists of polling place locations in 2002 and 2003. Matching and cleaning the files provided the bulk of the work. In the end, we have, for each voter who appears on both the 2002 and 2003 voting lists, the location of their polling place in each year, whether or not they voted in 2002, and whether or not they voted in 2003. In addition, we have other information of varying quality about people's sex, age, nationality and party registration. Appendix 1 describes the data matching project in more detail.

Our interest in this paper is in voters who had voted in Los Angeles County in 2002 and who were still on the registration rolls in 2003. We also wanted voters for whom a legitimate vote disposition had been recorded of either "V" for voting at the polling place, "A" for voting absentee, or "N" for not voting. Our consolidated 2002 and 2003 file had 4,172,149 individuals, but of these, 289,300 were new registrants in 2003 who were not on the file in 2002,[2] and 530,229 others dropped off between 2002 and

---

[2] We would expect that the impact of precinct consolidation on this group would be much less than on those who were registered in 2002, and we intend to do a separate analysis of their behavior.

2003. After excluding these people, 3,352,620 remained. We excluded 168,073 more

people who had moved between 2002 and 2003 based upon their addresses in our file.[3]

This left 3,184,547 people. Of these, 42,024 did not have a legitimate vote disposition on

the 2002 vote question and 5,630 more did not have a legitimate vote disposition on the

2003 vote question. These appear to be people who had actually registered after the date

of either the 2002 or the 2003 election. With these exclusions, the file contained

3,136,893 people. Finally, we excluded 25,826 more people in voting precincts for

which we did not have any information about their polling place location in either 2002

or 2003. Our final file has 3,111,067 voters – 74.6% of the number we started with, and

for the 25.4% of the individuals whom we dropped, 97.6% were not participants in one of

the two elections or they had moved so they did not fit the requirements for our analysis.

     For the 3,111,067 voters in our file, a change in polling place might have made a

difference because they were participants in both the 2002 and 2003 elections and they

had not moved. This is our basic analysis file for which we know whether a person's

polling changed and his or her vote disposition in both 2002 and 2003. We were able to

geocode the addresses of 3,045,206 of these people (97.8% of them), and we work with

that subsample when we include distance to the polling place in our analyses. One other

major variable that we use is age, and we have a reasonable birth year for 2,844031

people (91.4% of them) in our final file.[4]

---

[3] We would expect that the impact of precinct consolidation on this group would also be much less than on those who were registered in 2002 and preliminary analyses suggest that this is so.

[4] Some birth years are recorded as in the 1700's or 1800's. While there might be a few people over 102 in our file, we decided to exclude everyone older than this. (Another reason for doing this was that there was a very large number of people – an order of magnitude larger than 1899 or 1901-- with a recorded birth year of 1900.)

Precinct Consolidation in Los Angeles – In the 2002 general election there were 5,231 voting precincts in Los Angeles County. In the 2003 election, there were only 1,885 voting precincts. Roughly speaking, every three polling places had been reduced to just one. No surprisingly, doing this increased the average distance that people had to travel to their polling places. Figures 3 and 4 present histograms for the distance that people traveled to their polling places in 2002 and 2003 among the 3,045,206 for whom we have geocoded addresses. Note that there is a clear shift to the right in 2003 with the average distance going from .348 miles to .502 miles – a mean increase of .154 miles, or 44.3%. In addition, the percentiles of the distribution shifted as shown in Table 1, and the median went from .27 to .41 – a change of .14 miles, or 51.9%. In effect, the 25th percentile of voters in 2003 had to go the same distance as the 50th percentile in 2002, and the 75th percentile in 2003 looks had to go the same distance as the 90th percentile in 2002.

Initial Results – If we assume that the decision in 2003 about which polling places were changed was essentially random (an assumption that we will test in this paper), then the precinct consolidation constitutes a natural experiment for comparing what happens with and without changes in polling places. We can use those precincts whose polling place was unchanged to get estimates of P, A, and N, and we can use those precincts whose polling place was changed to get estimates of P*, A*, and N*. The first two rows of Table 2 provide these estimates, and the third row computes the differences between them. Note that there is an absolute 3.18% move away from polling place voting. But more than half of this, 1.72%, is made up by a turn to absentee ballots. Hence, the total

14

decline in turnout is only 1.45%. The reported t-statistics suggest highly significant results.

There are two reasons, however, to be wary of these results. First, assignment to treatment or control status was not by individuals; it was by 9,275 registration precincts which are pieces of election precincts,[5] and there many fewer (roughly one three hundredth) registration precincts than people. Obviously, even if consolidation were random with respect to these registration precincts, we cannot calculate statistical significance based upon the notion that people were randomly assigned to the treatment or control condition. Roughly speaking, the t-statistics assume a sample size of 3×10^6 units when they should assume something like 10^4 units. Hence, the standard errors from which the t-statistics are computed should be adjusted upwards by multiplying by the square root of the ratio of the 3,111,067 registrants to the 9275 registration precincts – by the square root of 335 which is about 18. When the standard errors are increased in this way, the t-statistics decline in the same ratio. In Table 3 we still have statistical significance when we do this for voting at the polling place and absentee voting, but not for non-voters. This "rough and ready" analysis suggests that we need a better way to calculate these standard errors.[6]

The second problem is that we have not proved that the consolidation was essentially random with respect to the potential outcomes for 2003 voting. One way to

---

[5] Each registration precinct was assigned to an election precinct. Each election precinct was composed of one or more registration precincts, and one's election precinct determined one's polling place. The number of registration precincts only changed by seven from 2002 to 2003 (from 9889 to 9882), and a small percentage of people's assignments to these registration precincts also changed, but the number of such precincts and the assignment of people to them is relatively stable. In our analysis file, we have 9275 registration precincts in 2003. Assignment in our file is basically by these 9275 registration precincts, not by the three million individual registrants of Los Angeles County.
[6] The analysis is "rough and ready" because the registration precincts vary substantially in size, thus complicating the adjustment.

get some good evidence about this is to consider the treatment's relationship to 2002

voting which is an excellent proxy for what might have happened in 2003 if there had not

been any consolidation. The bottom panel of Table 2 presents these data. The

differences between the treatment and control group are generally much smaller (one-

third to one-eleventh the size), but the t-statistics are all significant suggesting significant

pre-existing differences. Yet, the discussion above suggests that these t-statistics must be

approached skeptically. Indeed, if these t-statistics are all divided by 18 as suggested

above, they all fade into insignificance.

Thus, the first order of business must be to get better estimates of these t-statistics

than those presented in Table 2. These estimates were produced by standard ordinary

least squares regressions of the three dependent variables in Table 2 (*P, A,* or *N*) on a

dummy variable for whether a person's polling place was changed between 2002 and

2003. We must find a technique that takes into account the fact that randomization was

by people assembled together into registration precincts even though we have

observations on the individuals themselves.

This nested structure of data – discrete units inside of discrete units – is common

in social science, and one way to deal with it statistically is to use a hierarchical linear

model, HLM (Raudenbush and Bryk, 2002). HLM's effectively model both levels of

analysis and their attributes and compute the appropriate standard errors. To see how

HLM's work, we compare ordinary least squares methods with an HLM model. An OLS

method estimates the impact of changing polling places on a dependent variable such as

polling place voting by regressing an individual *i*'s polling place voting ($P_i=1$ if voting at

polling place, zero otherwise) on a dummy variable G_i for whether or not the polling

place was changed by grouping together precincts ($G_i=1$ for change, zero otherwise):

(7) $\qquad P_i = \beta_0 + \beta_1 G_i + \varepsilon_i$

The HLM estimation equation takes both individual registrants (i) and registration

precincts (j) into account by having both a "first-level" equation for individuals and a

"second-level" equation for registration precincts. In the first-level equation, individuals

are also identified by their registration precinct and the regression coefficients are

assumed to vary by registration precinct (hence the subscript j on both P and β):

(8) $\qquad P_{ij} = \beta_{0j} + \varepsilon_{ij}$

In the second level equation, the regression coefficient (in this simple model, only the

intercept β_{0j}) is assumed to vary from registration precinct to registration precinct based

upon the registration precinct's treatment status:

(9) $\qquad \beta_{0j} = \gamma_{00} + \gamma_{01} G_j + u_{0j}$

The inclusion of an error term u_{0j} also accounts for the fact that the treatment effect might

vary from one registration precinct to another.

     The first column of Table 3 reproduces the estimates from Table 2 for 2003 and

2002, and it adds some new estimates using HLM in the second column. The new HLM

estimates for 2003 are about 10-20% smaller than the earlier ones and the 2002 results

similar to the ones obtained before, but more importantly, the t-statistics (based upon

robust standard errors[7]) are smaller by factors ranging from one-half to one-fifteenth.

This is not quite as large a decline as we expected from a back-of-the-envelope

calculation, but it is substantial. The net result is that all the 2003 differences look quite

---

[7] The robust standard errors differ very little from the standard errors that are typically reported.

significant and the 2002 differences are insignificant for two out of the three dependent variables. The significance of the result for absentee voting in 2002 suggests that there might have been some pre-existing differences that should be taken into account.

One way to take into account preexisting differences in 2002 is to calculate change scores by subtracting the 2002 outcome for each person from the 2003 outcome. This process produces differences and t-statistics very similar to the 2003 results except that the decline in voting (using the HLM estimates) increases from 1.28% to 1.61% because absentee voting takes up less of the reduction in polling place voting (only 1.00% rather than 1.40%).[8] The results for absentee voting reflect the adjustment for the overrepresentation of absentee voters in consolidated precincts.

Table 4 takes into account past voting behavior in 2002, since that is a powerful predictor of one's voting behavior in 2003. In each case, despite the powerful tendency toward inertia, a change in one's polling place had strong substantive effects on voting behavior. We again present both OLS and HLM estimations and their accompanying t-statistics. The pattern continues to obtain in all cases – the absolute values of the coefficients decrease slightly and the t-statisticss typically drop by about two-thirds when estimating by HLM rather than OLS. We will only discuss the HLM results here.

While behavioral changes in the expected directions are observed in all cases, the largest impact is on voters who voted at the polls in 2002. It seems likely that these people would be most affected by a polling place change since they had gone to their polling place in 2002. Among those who voted at the polling place in 2002, those whose

---

[8] This effect follows from the obvious fact that the differences in the corresponding rows in the 2003 and 2002 parts of the table must equal the results in the last three rows of the table. Thus, using the HLM estimates, the 1.40% difference in absentee voting in 2003 minus the 0.40% pre-existing difference in 2002 absentee voting equals 1.00% which is the difference in absentee voting reported in the second of the last three rows of the table.

polling place changed voted in person 4.22% less in 2003 than those whose polling place did not change. And a substantial majority of these people chose to not vote rather than switch to absentee voting (55.9%), which is unsurprising since we can infer that most of these people do not habitually vote absentee.

Among people who voted absentee in 2002, those whose polling places were consolidated increased their absentee voting by 0.78 percentage points more than those whose polling places remained the same. These extra absentee voters in the consolidated group were drawn from the 1.08% less who voted at the polling place. Among those registrants who did not vote in 2002, 1.71% fewer went to the polling place in the consolidated group than in the unconsolidated group. Of that decrease, a majority was repeat non-voters (63.2%), while the rest resorted to absentee balloting.

The last column of Table 4 calculates the relative impact for each group by taking the percentage change over the percentage in the population in 2002 in each group. These numbers are multiplied by 100 to make them into relative percentages. This statistic adjusts the absolute percentages in the first and second columns by the differences in the sizes of the polling place, absentee, and non-voting groups. Relatively speaking, those who voted in their polling places or by absentee in 2002 are the most likely to decrease their polling place voting (by -10.14% and -10.09% respectively) in 2003 if their polling place is changed. Not surprisingly, those who did not vote in 2002 are much less likely to be affected by such changes – because their most likely action is to not vote at all whether polling places are consolidated or not. It is also not surprising that those who voted absentee in 2002 are much more likely to adjust for a change in

their polling place by choosing to vote absentee in 2003 (7.29% of 2002 absentee voters versus 4.50% of 2002 polling place voters and 1.30% of non-voters.)

Impacts of Age and Distance to the Polling Place – These preliminary results suggest that there were significant impacts on voting from the consolidation of polling places, although they also suggest that we might have to pay more attention to pre-existing differences before we can be sure of this result. We can get a better fix on the need to do this, while also learning about the determinants of voting, by considering some other factors that we suspect are strongly related to voting and that might explain pre-existing differences. In this section we consider distance to the polling place and age. If assignment is essentially random, then we would expect there to be no significant differences in the characteristics of the treatment and control groups. If the characteristics of the treatment and control groups do differ, however, we might expect that by conditioning on those characteristics that pre-existing differences in the two groups as measured by voting behavior in 2002, for example, would disappear. In effect, we would be controlling for those differences. If there are significant differences in characteristics and pre-existing differences do not disappear when we condition on each one alone, then we must take a multivariate approach to adjusting the treatment and control groups to ensure comparability. We might also have to think about other characteristics that might explain the pre-existing differences.

Consider distance to the polling place in 2002. Those who had their polling place changed in 2003 had to go an average distance in 2002 of .354 miles whereas those who did not have their polling place changed had to go only .321 miles—a difference of .034 miles which is highly statistically significant. Perhaps this explains the pre-existing

differences in voting. For this to be true, there has to be some relationship between distance to the polling place and the likelihood of voting.

Figure 5 shows a graph of polling place turnout in 2003 by the distance, in hundredths of a mile, a voter had to travel to get to the polls in 2003. We distinguish those voters who lived in consolidated voting precincts from those who live in unconsolidated voting precincts, and we know the distance each voter had to travel to the polls for about thirty-nine of every forty voters. It is clear that there is a huge difference here – people who were able to go to the same polling place in 2003 as they did in 2002 consistently vote at higher rates than those who were assigned to a new polling place in 2003. In addition, there is a somewhat noisy but clear negative relationship between distance to the polling place and the likelihood of voting for both the treatment and control group. Perhaps distance to the polling place explains the differences we found earlier in 2002 voting between the two groups, and perhaps we can condition on distance to the polling place in 2002 to provide further evidence of real differences between the two groups.

Strong evidence for real differences comes from Figures 6 and 7 which plot polling place voting in 2003 versus 2002 distance to the polling place and polling place voting in 2002 versus 2002 distance to the polling place. Figure 6 demonstrates a real difference between the groups in 2003 while Figure 7 suggests that there was no pre-existing difference in polling place voting. In Figure 6 the lines for the treatment and control only cross twice, but the two lines cross 34 times in Figure 7 which is roughly half the 73 categories.[9] This strongly suggests that there were no significant pre-existing

---

[9] It is easy to see that if each treatment category for a given distance category has the same Bernoulli probability for polling place voting (a null hypothesis of no difference between the two lines), then the

differences in polling place voting with respect to 2002 distance to the polling place. It also strongly suggests that the observed differences in 2003 are highly significant.

Figure 6 also suggests one other interesting fact. While there is a steady decline in turnout among those who voted in the same place in 2003 as distance increases, the turnout rate among those whose polling place changed, while lower, is fairly steady as distance increases, hovering around 44-45%. This suggests that something about the 2003 treatment not only decreased polling place voting for everybody, no matter how far they were from their polling place in 2002, but it also counteracted the effects of distance from the polling place that is evident for those who did not have their polling place changed. Figure 8 suggests what happened by plotting the change in distance to polling place between 2002 and 2003 versus the distance to polling place in 2002. Clearly, the consolidation had the dual effect of increasing the average distance to the polls while making the distances more equal. Thus, those people within one-hundredth of a mile of the polling place in 2002 had the distance to their polling place increased by six-tenths of a mile, while those within about .65 miles in 2002 had no increase in the distance to their polling place. And those beyond .65 miles actually had the distance to their polling place shortened from what it was in 2002. Thus, the near constancy of polling place voting in Figure 6 among those whose polling place was changed appears to be the result of this variation in the treatment which compensated for some of the handicap that those far from the polling place had traditionally experienced. Moreover, Figure 5 shows that

likelihood that the one estimate will be above another should be one-half. Furthermore, adjacent estimates (in terms of the distance) for the same treatment should be both above or both below the estimates for the other treatment exactly one-half the time. Hence, we should see crossings about half the time if the two lines are from the same populations. Using a binomial test, it is easy to see that thirty-four crossings out of 72 possibilities (there is one fewer crossing possibility than categories) could easily occur under the null hypothesis whereas two out of 72 crossings (as in Figure 5) is highly improbable.

there was still a negative relationship between distance to the polling place and polling place voting in 2003 even among those who had their polling place changed.

We now consider absentee voting in 2003 and 2002 by distance to the polling place in 2002. In Figure 9 we find the expected reciprocal effect to the 2003 polling place turnout. A wide gap exists between the two groups; this time, however, those people whose polling places were unchanged resorted to absentee voting at lower rates than did people who faced a new polling place. As this additional information cost was added to the cost of voting at the polls, voting absentee became more attractive. As 2002 distance increases, absentee turnout rate increases for both. The lines begin to cross at distances exceeding five miles and track very closely after that threshold.

Once again, however, there may be pre-existing differences that explain these results. When we look at absentee voting in 2002 by 2002 distance to the polling place in Figure 10, we see much less spread between the two lines that in Figure 9, and the two lines do cross 20 times,[10] but there also appears to be a slight gap with more absentee voting in 2002 among those who had their polling place changed. Some preexisting differences do exist between the people in precincts which were consolidated and precincts that were not with respect to absentee voting, and conditioning on distance to the polling place in 2002 does not make these differences go away. This is not surprising – it makes sense to consolidate districts with a higher proportion of absentee voting, because the change in the physical polling place location will affect a lower percentage of voters in those precincts. Nevertheless, the difference between the groups is much less in 2002 than in 2003, so while we must condition and adjust for those preexisting

---

[10] Not quite enough crossings to conclude that the two lines are identical; we need at least 25.

differences in absentee voting, those alone do not explain the whole of the group differences in 2003 in absentee voter turnout.

We conclude that polling place distance does have an impact on voting and that 2002 polling place distance does differ across the two groups. Conditioning on it might help to reduce differences between the two groups, but it does not make the differences in absentee voting in 2002 go away completely.

Another important determinant of voting is age, and we have a reasonable year of birth for about 91.4% of the voters on the file. There are pre-existing differences between the two groups in age. Most importantly, there are about .73% more people over age 60 in the group that did not have its polling place changed. Since the elderly are more likely to vote absentee, this might explain the pre-existing differences in 2002 absentee voting.

Figure 11 plots polling place vote by treatment for ages between 20 and 90 – those for which we have large enough numbers of people to make relatively smooth plots. Clearly there appear to be significant differences between the two treatment groups. This figure also displays the well-known inverted U of voting which increases with age until the mid-fifties and then slowly declines to low levels by the 80's. In addition, it provides a very strong suggestion that polling place voting may have decreased more among older people than younger people.

Again, however, these differences might be pre-existing so Figure 12 plots polling place voting in 2002 by age. Although the treatment and control lines track very well at younger ages below 55, above that age they begin to diverge and there is higher polling place voting in 2002 among those who did not have their polling place changed in 2003.

Thus, controlling for age actually reveals a pre-existing difference in polling place voting among the elderly. It seems likely that special care might have been taken to not change polling places for those senior citizens who had voted at the polling place in 2002. Indeed, we have been told by Los Angeles county officials that they tried to avoid consolidating precincts with high proportions of older people who voted at the polling place.

Figure 13 plots absentee voting in 2003 by age, and the two lines reproduce the well-known result that absentee voting increases with age. In addition, we see the same steady gap between the two lines suggesting that some voters dissuaded from voting at the polls by a change in polling place location resort to absentee voting. Moreover, the impact on absentee voting of changing polling appears to be greatest among older people. Figure 14, however, suggests that at least some of the increased absentee voting for older people as a result of changes in polling places might have been pre-existing. This appears to be another manifestation of Los Angeles County's attempt to consolidate precincts in ways that mitigated the impact of consolidation.

The Consolidation Process and Its Consequences

The preceding discussion suggests that at least three different factors affected the assignment process in Los Angeles County: the likelihood of voting absentee, the 2002 distance to the polling place, and the age of the person. Tables 2 and 3 provide evidence about greater absentee voting in 2002 among registration precincts that had their polling place changed. Figures 15 and 16 provide this demonstration for 2002 distance to the

polling place and age. Figure 15 shows that as distance increased, the likelihood that a registrant's polling place would be changed increased from less than 65% to over 70%. Figure 16 shows that as a person's age increased, the likelihood that his or her precinct would be changed dropped from over 66% to the low sixties.

The Assignment Process -- These results indicate that we must take into account the selection or assignment process that assigned registration precincts to election precincts, thus either changing the polling place of those in the registration precinct to new location in 2003 or keeping the polling place in the same location. In truth, it is clearly far too optimistic to believe that assignment would have been completely random because it makes no sense to "consolidate" precincts by randomly linking one precinct with another – the result could be the conjunction of precincts miles and miles away from one another with no common boundary. A better way to consolidate precincts is to take a group of contiguous precincts and to join them into a larger precinct. Then some polling location in a precinct near the center of agglomeration is chosen as the polling location for the consolidated precinct. Because adjoining precincts are probably similar in many ways this process amounts to "matching" or "pre-stratifying" precincts, and then choosing the polling location for the consolidated precinct from one of the matched precincts. If the polling location were chosen randomly from those available in the matched precincts, then this would be a good randomized experiment with some stratification that might substantially improve statistical efficiency. But the polling location is probably not chosen randomly, it is probably chosen to be near the middle of the matched precincts. Nevertheless, it seems likely that this amounts to a relatively random procedure for choosing polling locations. In fact, it is essentially the geographic

analog of classic random sampling procedures from lists in which a random starting place was chosen and then every tenth name was added to the sample.

Another factor, however, might diminish the randomness of this procedure. Decision-makers might have taken some characteristics of the voters or the precincts into account when they consolidated them. If these characteristics affect turnout in a significant way, then there could be substantial differences between the voters with changed and unchanged polling locations. In the worst case scenario, a highly partisan Registrar of Voters might change polling locations for precincts with large concentrations of partisans of one particular party. This kind of machination might occur in some places (Florida, Illinois, and Texas) with highly partisan County Registrars of Voters, but it seems unlikely in California with its Progressive tradition of choosing Registrars based upon their efficiency and effectiveness and not their partisan identification. In this case, it seems much more likely that a Registrar of voters would consolidate precincts in order to mitigate some of the reduction in turnout that might be expected from such a consolidation. To do this, the Registrar would have to have some theory about what affects voting turnout and would have to have some data on the precincts. The most available data would be from the voter registration file, and the theories would presumably be fairly simple – such as believing that past voting behavior predicts future voting behavior, that travel distance matters, and that age affects voting behavior. This simplifies the modeling process because it suggests that we probably have at hand most of the information that was used by the Registrar in the consolidation process.

Specifically, we would probably expect the Registrar to consolidate precincts with the following factors in mind. First, we would expect to find that registration precincts

with greater fractions of absentee voters would have their polling place changed – after all, for absentee voters, changing their polling place has no impact on their turnout. Second, we would expect to find that registration precincts with more elderly voters would be less likely to have their polling place changed (especially if they had large numbers of polling place voters). Third, we would expect that smaller registration precincts would be more likely to have their polling place changed than larger ones because fewer voters would be discomfited by changing the location of a polling place for a small registration precinct. Finally, we might expect that the consolidation decision would take into account 2002 distance to the polling place for the voters in a registration precinct. The empirical data we have presented suggests that the Registrar tried to reorganize polling places that were on average far away from voters in such a way as to minimize increases in distance. This presumably required changing more of these polling places.

Following the classic works by Heckman (1979) and Achen (1986), the selection process is modeled by regressing a binary variable for "treatment" or "control" on the covariates thought to be important for the selection process. Because the decision was made by registration precincts, we use these as the unit of analysis. Hence, those variables related to individual characteristics or the distance to the polling place are averages. In addition, we must deal with a peculiarity of the registration precincts. It was possible (but unlikely) for individuals to be moved from one registration precinct to another between 2002 and 2003. Consequently, even though all members of a registration precinct are always assigned the same polling place location, it is possible for a registration precinct for which almost everyone had their polling place changed to still

have some people for which their polling place was not changed. The reverse is also possible. Figures 17 and 18 show, however, that it is very unlikely. Almost all the registration precincts (95% of them with 95% of the people in them) are entirely composed of either people who did not have their polling place changed (at zero in the histograms) or who did have their polling place changed (at 100 percent in the histograms), as seen in Figure 17. Nevertheless, there are some "mixed" registration precincts as shown in the "blowup" in Figure 18. In order to accommodate them, we dichotomize the variable for the percentage of people who changed their polling place at one-half. Finally, to simplify the process of interpreting the regression, we present results from a linear regression (or a linear probability model without a Goldberger correction) although a logistic or probit yields exactly the same results.

Table 5 presents the results for two regressions of the treatment variable on the four factors that we thought would be important for assignment and some others which we thought would not be. The size of the registration mattes and larger ones are less likely to have their polling place changed. Thus a relatively large registration precinct of 500 people (at the 75th percentile in size) is about five percent less likely to have its polling place changed than a small registration precinct of 100 people (at the 25th percentile). And a registration precinct with a relatively high level of absentee voting at 12 percent (at the 75th percentile) is about two percent more likely to have its polling place changed than a registration precinct with only 6% absentee (at the 25th percentile). And a registration precinct in which its voters have to go one-quarter of a mile more to get to their polling place has a four percent greater chance of having its polling place changed. Finally a registration precinct that goes from 14 percent over age 60 (the 25th

percentile) to 24 percent over age 60 (the 75th percentile) decreases its chances of having its polling place changed by almost two percent.

In addition, the fraction of people age 40 to 59, the fraction of polling place voters, or the fraction of Democrats, Republicans, or Independents does not affect the selection probability. The second column of regression coefficients adds one additional fact – it is not the percent of those over age 60 which matters so much as its interaction with the percent who vote at the polling place.[11]

Despite these findings, perhaps the most important thing about this regression is that it explains very little of the selection process – the R-squared is only .012 meaning that only about 1.2% of the variance is explained. Another important thing is that the selection process certainly seems to follow strategies to mitigate the adverse impacts of consolidation. This means that we should seek an approach that will allow us to correct for these mitigating strategies.

Correcting for Non-Random Assignment – There are a number of ways to correct for non-random assignment such as selection adjustment in regression, matching, and propensity score methods (Rosenbaum 2002; Imbens 2003). A selection adjustment would proceed by estimating the selection equation as above for registration precincts j where G_j is one if the registration precinct is treated and zero otherwise:

(10) $\quad G_j = \alpha_0 + \alpha_1(\%\text{ Age Over }60)_j + \alpha_2(\text{Average Distance to Polling Place 2002})_j +$

$\qquad\qquad \alpha_3(\%\text{ Absentee Vote in 2002})_j + \alpha_4(\text{Size of Registration Precinct})_j + \varepsilon_j$

---

[11] This interaction suggests that we should look more carefully for other interactions that follow from sophisticated decision rules by the County Registrar.

Then we would estimate an outcome equation for individuals such as the following for polling place voting (P_{ij}) where G_{ij} associates individual i with the proper registration precinct:

(11) $P_{ij} = \beta_0 + \beta_1 G_{ij} + \beta_2 \text{(Age)}_i + \beta_2 \text{(Distance to Polling Place 2002)}_i$

$+ \beta_3 \text{(Absentee Voter 2002)}_i + \delta_i$

Note that we have assumed that the size of the registration precinct does not affect the outcome process. Theoretically it makes no sense for it to do so, although it is possible that it is a proxy for other factors that should have been included in the specification. Since there might be some unobserved characteristic that affects both the selection process and the outcome, the error terms ε_j and δ_i might be correlated, and an ordinary least squares estimate of the coefficients in the outcome equation will be biased. Specifically, the estimate of the treatment, β_1, will be biased. However, the size of precinct can be used as an instrumental variable in the second equation. (One might also be able to use the fact that the remaining variables in the first equation are aggregated for precincts in order to use them as instruments in the second equation, but this approach seems fraught with difficulties.)

This approach relies heavily upon trusting an instrument and believing a linear specification. It also confronts some additional problems because the selection process occurs on a different level from the outcome. We have also found that the selection equation does not explain much variance so we should especially wary of its specification. Indeed, we know that there may be complicated interactions among the

variables as shown with the interaction of the elderly and polling place voting. It would

be better to use a method that avoided these problems.

Matching seems like a suitable approach. Matching deals with the fundamental

problem of causal inference, the fact that we can only observe each unit in either the

treated or control condition, but not both, by taking each observed unit and matching it

with one or more other similar observed units that received the opposite treatment. Thus,

for a unit that received the treatment, we find a unit with similar characteristics that did

not receive the treatment. We then use this unit as the imputed control for the treated

unit. The trick, of course, is to match on the right characteristics, and this must be done

to satisfy what is called the "conditional independence assumption" or the

"unconfoundedness" assumption. To state it properly, we use the potential outcome

notation that dates back to Neyman (1923) and that was first explicated in detail by Rubin

(1974) and Holland (1986). In this notation, each outcome, such as polling place voting,

P_i, is given an argument representing its value if it had received the control status $P_i(0)$ or

the treatment status $P_i(1)$.   Only one of these is actually observed. Then the requirement

of the unconfoundedness assumption is that conditional on a set of exogenous

characteristics X_i, both $P_i(0)$ and $P_i(1)$ are independent of the treatment. This assumption

fails, for example, if the treated group would have had higher voting at the polling place

if it had not been treated than the untreated group. That is, if we represent the group of

people who got the treatment by the subscript t and there are T of them, while the people

who are in the control group are represented by the subscript c and there are C of them,

then the assumption fails if $P_T(0) = \sum_t P_t(0)/T$ is greater than $P_C(0) = \sum_c P_c(0)/C$. Thus,

if we think of the vote in 2002 as a good surrogate for the potential outcome if

consolidation had not occurred, then the fact that the mean absentee vote in 2002 for the group that is treated in 2003 is higher than the mean absentee vote in 2002 for the group that is in control status in 2003 indicates that this assumption fails.

The trick, then, is to condition on variables that create pre-existing differences and that could affect the outcomes. The obvious candidates are the four that we had in our selection equation, but, in fact, if size of precinct really does not affect the outcome, then it can be dropped from the list.[12] Thus, we have matched cases on voting behavior in 2002, age, and distance to the polling place in 2002. To accomplish this, for every observation for which we had sufficient data and for which a match was possible (2,722,240 or 87.5% of our analysis file), we found exact matches on age and 2002 voting behavior.[13] For distance to the polling place in 2002, an exact match required that the matched observation have a distance within plus or minus .01 miles. In the results reported here, we searched for up to ten potential matches and we used all that we found for each case. We then took the difference between the original unit's value on the dependent variable and the average of its matches. To get summary results, we averaged over both treatment and control units to get a sample average treatment effect.

Table 6 presents the results of this matching for our three outcome variables. The first column reports the raw results without matching for the 2,722,240 people who we were able to match. Except for the smaller sample size, these results are the same as the 2003 results reported in Table 2 and in the first column of Table 3. The second column reports results from using change scores (e.g., the difference between 2003 polling place voting and 2002 polling place voting) as reported in the last three lines of Table 3. The

---

[12] In future work, we will match on this variable as well, just to be safe.
[13] Almost all the cases for which no matches were found lacked either age or distance to polling place in 2002.

third column presents the matching results. The most notable conclusion is that non-voting appears to be much higher than in any other estimate. But we might have expected this since the matching tries to adjust for the mitigating efforts taken by the Registrar of Voters. In fact, we can think of the matching results for non-voting as what would have happened had the Registrar not exerted any efforts to mitigate the impact of consolidation, and the raw results as what did happen. The difference is relatively small – only .26 percent – suggesting the difficulty of overcoming the problems of consolidating precincts.

These results are presented without standard errors. In future work, we intend to use methods such as those suggested by Imbens and his co-authors (Abadie, Drukker, Herr, and Imbens 2001, Abadie and Imbens 2004) to estimate standard errors. Rather than pursuing this approach here, we turn to estimating a structural model of voting turnout.

Multinomial Model

Earlier in this paper we presented a model of how the costs imposed on individual voters by changing polling places could affect individual voting, and we argued that this model could help us to learn about the costs of voting. In this section, we estimate a structural model of voting which reveals the relative importance of some cost factors, such as information and transportation costs, that we identified earlier. To do this we must modify the theoretical model presented earlier.

Theory -- In the earlier model, we wrote individual utilities U_i^m for polling place ($m=p$), absentee ($m=a$), and not voting ($m=n$) in terms of benefits and costs b_i, c_i^p and c_i^a.

For example, we wrote the utility for polling place voting as: $U_i^p = b_i - c_i^p$. If we observed enough information about each individual then we would be able to calculate exactly these benefits and costs for voting, but we can only observe some of the factors that affect individual benefits and costs so that we have to partition people's utilities U_i^m into those benefits and costs that we can observe plus some stochastic term representing those benefits and costs that we cannot observe, that is $U_i^m = V_i^m + \varepsilon_i^m$ where V_i^m consists of observed benefits and costs and ε_i^m represents unobserved benefits and costs. To relate this model with unobservable cost and benefits to our earlier one without them, we should partition benefits and costs (b_i, c_i^p, and c_i^a) into observable and unobservable ones, but rather than add more notation, we assume that <u>observed</u> benefits and costs are denoted by b_i, c_i^p, and c_i^a, although a more fastidious notation would add some mark to indicate that these are only the observed components of benefits and costs. Since these are observed benefits and costs, we set them equal to the observed component V_i^m of utility so that:

$$V_i^p = b_i - c_i^p.$$

(12) $V_i^a = b_i - c_i^a$

$$V_i^n = 0.$$

We then write the observed benefits and costs in terms of observed vectors of characteristics X_i, Y_i, and Z_i and unknown parameters α, β, and γ:

$$b_i = X_i\alpha$$

(13) $c_i^p = Y_i\beta$

$$c_i^a = Z_i\gamma$$

And using this notation we can then write the fixed part (V_i^p, V_i^a, and V_i^n) and the

stochastic part (ε_i^p, ε_i^a, and ε_i^n) of people's utilities as follows:

$$U_i^p = V_i^p + \varepsilon_i^p = (b_i - c_i^p) + \varepsilon_i^p = (X_i\alpha - Y_i\beta) + \varepsilon_i^p$$

$$(14) \quad U_i^a = V_i^a + \varepsilon_i^a = (b_i - c_i^a) + \varepsilon_i^a = (X_i\alpha - Z_i\gamma) + \varepsilon_i^a$$

$$U_i^n = V_i^n + \varepsilon_i^n = 0 + \varepsilon_i^n$$

If we assume that the stochastic terms ε_i^m are distributed independently with a Weibull

distribution, then we can express choice probabilities as the multinomial conditional

logit:

$$(15) \quad \text{Prob}(m=1) = \exp(V_i^m)/\textstyle\sum_m[\exp(V_i^m)]$$

So that, for example, the log odds of choosing to vote at the polling place versus not

voting is the following:

$$(16) \quad \ln[\text{Prob}(p=1)/\text{Prob}(n=1)] = \ln[\exp(V_i^p)/\exp(V_i^n)] = V_i^p = (X_i\alpha - Y_i\beta)$$

Where the next to the last equality comes from remembering that $V_i^n = 0$ and the last one

from using (12) and (13) above.   And similarly:

$$(17) \quad \ln[\text{Prob}(a=1)/\text{Prob}(n=1)] = V_i^a = (X_i\alpha - Z_i\gamma)$$

With these expressions for the log odds and with data on individual decisions to vote and

individual characteristics X, Y, and Z, we can estimate the parameters α, β, and γ using

methods such as maximum likelihood (McFadden 1973).

   Determinants of Voting Turnout -- The crucial remaining step is to define the

variables comprising vectors X, Y, and Z such that they capture the major benefits and

costs so that what remains in the stochastic terms ε_i^m is idiosyncratic variation that is

uncorrelated with these variables. At the moment, we are still working to add Census

block and tract data to each individual observation which will, among other things,

provide information on the level of education in the Census block in which each person lives. These data will make for a better model, but for the moment we must use what we have. What we have are the following variables which we have organized into benefits, costs of polling place voting, and costs of absentee voting:

Benefits:

Major Party Registration – Registration with a major party is an indicator of interest in the political system and in the elections which primarily involve the two major parties. Using the voter file, we can determine whether a registered voter is affiliated with either the Democrats or Republicans.

Vote in 2002 – Voting either in the polling place or by absentee indicates interest in the political system.

Age and Age-squared – Other research has shown that interest in politics increases with age, especially from age 20 to age 40, although it eventually levels off and may even decline.

Costs of Polling Place Voting

Distance to Polls in 2002 – Theory suggests that because of transportation and other costs, increased distance to the polls should increase the cost of polling place voting, and the bivariate plots presented earlier suggest such a relationship.

Vote at Polling Place in 2002 – Voting at the polling place in 2002 suggests that you are a person for whom the costs of polling place voting are relatively low.

Age and Age-Squared – Mobility and safety concerns increase with age, especially past age 50, so we might expect that increasing age would mean increased costs of polling place voting.

Change in Polling Place – The fact that polling places were changed in a way that amounts to a natural experiments means that this is a very good proxy for the costs of having a polling place changed. As noted earlier, these costs are related to information, transportation, and risk aversion. By including the change in distance to the polling place (see below), the dummy variable for the change in polling place provides an estimate of the impacts of information and risk aversion.

Change in Distance to Polling Place – We would expect that an increase in the distance to the polling place between 2002 and 2003 would increase the costs of polling place voting.

Costs of Absentee Voting

Permanent Absentee – In Los Angeles County, a voter can request that he or she become a permanent absentee so that the documents for absentee voting are simply sent out every election. Obviously this would decrease the cost of absentee voting.

Vote Absentee in 2002 – Voting absentee in 2002 obviously suggests that the costs of voting at the polling place are too high for an individual.

This discussion leads to specific expectations for the signs of each variable in the multinomial logit link equations (16 and 17 above), and it leads to some exclusion restrictions from each of the two equations.

Estimation of the Multinomial – We used the multinomial program in HLM, version 5.05, to estimate this model. Unfortunately, the HLM multinomial requires that the same variables be in both level one equations for explaining polling place voting (relative to not voting) and absentee voting (relative to not voting). Consequently, we were not able to impose our theoretically derived restrictions on these variables. We indicate these variables by asterisks in the results reported in Table 7. We estimated our model in two ways that make different assumptions about how distances affect voting. In the first model called "Distance Grand-Centered" we simply included the distances in the model. In the second model called "Distance Centered by Registration Precinct" we assumed that relative distances matter for people and that their peer group was other people in their registration precinct.

The results are generally in accord with our expectations. Almost all the signs are correct, except for 2002 distance to polling place in the "Distance Grand-Centered" model for polling place voting. But the basic treatment variables, in bold, have the right sign and exert a robust impact. In the first model, the variables that we thought should be excluded do no appear to be insignificant, suggesting a misspecification of our model (leaving out education is surely a problem). But in the second model using relative distances, at least one variable that we thought would be zero in the absentee voting model, the change in distance to polling place between 2002 and 2003 goes to insignificance, and the anomalous sign on 2002 distance to the polling place in the polling place equation is reversed.

The size of the treatment effects is interesting. The dummy variable for the change in polling place has a coefficient that is roughly the same size as the coefficient

for the change in distance, but whereas the treatment effect for the dummy variable goes from zero to one, the typical treatment effect for the change in distance is about .15 miles—about a sixth of a mile. Hence, the disruption effect from simply having a change location is substantially greater than the transportation effect from an increased distance to the polling place.

Conclusions

Although the 2003 consolidation of precincts is not a perfect natural experiment, it is about as close as we normally come with observational data. Consequently, it provides us with a very strong inference that consolidation in Los Angeles County reduced turnout by a substantial 1.88% in the precincts in which the polling location was changed. We also find that voting at the polling place decreases even more, by 3.05% but that an increase of absentee voting of 1.19% makes up for some of this reduction. In addition, we find that the substitution of absentee voting for a reduction in polling place voting is greatest among people of middle age and older whereas younger people are more inclined simply not vote at all.

We also find that the change in polling place location has two effects: a transportation effect resulting from the change in distance to the polling place and a disruption effect resulting from the information required to find the new polling place and the risk aversion that people feel about going into a new neighborhood. The disruption effect is about five times larger than the transportation effect for the average person who experienced an increased distance to the polling place of about a sixth of a mile, but the effects were roughly equal for someone who had an increased distance of about a mile.

From a technical perspective, future work involves calculating standard errors for the matching and improving the estimation of the multinomial. From a substantive perspective, adding neighborhood level data will make it more possible to disentangle the reasons for why people are discouraged from voting when their precinct polling place is changed. It should also make it possible to estimate a cost function for voters.

Appendix 1

All the data acquired came from the Los Angeles County Registrar-Recorder/County Clerk. The voter data were obtained routinely; the polling place data less so. Individual-level voter data includes a voter's name, registration precinct, residential address, mailing address, phone, party registration, sex, birth date, birth place, date of registration, date of last transaction, permanent absentee voter status, and turnout records (in person voting, absentee voting, or abstention), along with several fields of identifying information and miscellaneous other data. Some of these data are incomplete; for example, dates and places of birth are missing in many cases, and sex is missing more often than not, although it can be inferred from the title field (Mr., Mrs., Miss) as well in some cases. Still others are obviously incorrect, such as an improbably large cohort of people born in 1900, as well as even "older" people born in the 18th and 19th centuries; we made a decision to code any birth date prior to the year 1901 as missing. However, the critical data of precinct and turnout are always present.

The Registrar only maintains official records of polling places in hard copy. These records include polling place precincts, polling place addresses, and a description of the polling place (residence, business, church, school, etc). While 2003 turned out to be available electronically via a stroke of luck, 2002 had to be scanned in using OCR software, and then reviewed line-by-line for correctness. We are grateful for the assistance of several colleagues at UC DATA and the Survey Research Center in executing this technically challenging and labor intensive task, including Ilona Einowski, Jon Stiles, Eva Seto, Lyn Civitello, Ricardo Gutierrez, and Virginia Nee.

We then had to match each voter to their polling place for both 2002 and 2003. This was a challenge, because the precincts reported for the voters were at a different level of precision then those reported for the polling place. Voters were associated with their registration precincts. Polling places were associated with their polling place precinct, which is composed of one or more (often many more) registration precincts. We needed to acquire "crosswalk" data to merge the two files so that each voter could be associated with a polling place precinct and its corresponding address. We obtained these data from both the Los Angeles County Registrar itself and from Karin McDonald of the Statewide Database at the Institute of Governmental Studies; we are grateful to both.

Having created complete files for 2002 and 2003, the final step was to match voters from the two years and look at voting behavior changes between 2002 and 2003, contingent upon whether one's polling place was moved or not. We used the unique identifier Voter ID to match voters from both years; approximately 3% were not matched, probably because of normal churning (residential moving, mortality, etc.) in the electorate.

References

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. (2001), "Implementing Matching Estimators for Average Treatment Effects in Stata," The Stata Journal, 1:1-18.

Abadie, A,. and Imbens, G. (2004), "Large Sample Properties of Matching Estimators for Average Treatment Effects," Working Paper.

Achen, C. (1986), The Statistical Analysis of Quasi-Experiments, University of California Press.

Campbell, A., Converse, P., Miller, W., & Stokes, D. (1960). *The American Voter.* New York: Wiley.

Downs, A. (1957). *An economic theory of democracy.* New York: Harper and Row.

Gimpel, J. G. & Schuknecht, J.E. (2003). Political Participation and the Accessibility of the Ballot Box. *Political Geography,* 22, 471-488.

Green, D. P. & Shapiro I. (1994). *Pathologies of rational choice theory: A critique of applications in political science.* New Haven: Yale University Press.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," Econometrica, 47:147-161.

Highton, B. (1997). Easy registration and voter turnout. *Journal of Politics.* 59(2), 565-575.

Holland, P. (1986), "Statistics and Causal Inference, Journal of the American Statistical Association, Vol. 81, No. 396. pp. 945-960.

Imbens, G. (2003), "Semiparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," Working Paper.

Knack, S. (1995). Does 'motor voter' work? Evidence from state-level data. *Journal of Politics,* 57(3), 796-811.

McFadden, D. (1973), "Conditional Logit Analysis of Quantitative Choice Behavior," in Frontiers of Econometrics, edited by Paul Zarambka, New York: Academic Press.

Miller, W. E., & Shanks, J. M. (1995). *The new American voter.* Cambridge: Harvard University Press.

Neyman, J. (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, Vol. 5, No. 4. (Nov., 1990), pp. 465-472.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd edition*. Thousand Oaks, CA: Sage Publications.

Riker, W. H., & Ordeshook, P. C. (1973). *An introduction to positive political theory*. Englewood Cliffs, NJ: Prentice Hall.

Rosenbaum, P. (2002), Observational Studies (2nd Edition), New York: Springer-Verlag.

Rosenstone, S. J., & Hansen, J. M. (1993). *Mobilization, participation and democracy in America.* New York: Macmillan.

Rosenstone, S. J., & Wolfinger, R. The effect of voter registration laws on voter turnout. *American Political Science Review*. 72(1), 22-45.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," Journal of Educational Psychology, 66: 688-701.

Squire, P., Wolfinger, R., & Glass, D. P. (1987). Residential mobility and voter turnout. *American Political Science Review, 81*(1), 45-66.

Stein, R. M. (1998). Early voting. *Public Opinion Quarterly, 62*(1), 57-59.

Wolfinger, R. E., & Rosenstone, S. J. (1980). *Who votes?* New Haven: Yale University Press.

TABLE 1:
Average Distance to Polling Place in 2002 and 2003

	Miles to Polling Place for Percentiles of Voters				
	10%	25%	50%	75%	90%
2002	.10	.17	.27	.43	.64
2003	.15	.25	.41	.63	.90

TABLE 2:
Raw Results for 2003 & 2002 by Polling Place Change

	Voted at Polling Place	*Voted Absentee*	*No Vote*	*Totals*
2003				
Changed Polling Place	910,384 44.327%	313,250 15.252%	830,141 40.420%	2,053,775 66.0%
Unchanged Polling Place	502,280 47.506%	143,032 13.528%	411,980 38.966%	1,057,292 34.0%
Total People	1,412,664 45.408%	456,282 14.666%	1,242,121 39.926%	3,111,067 100%
Difference	-3.179%	1.724%	1.454%	
t-statistic	-53.37	40.73	24.82	
	Voted at Polling Place	*Voted Absentee*	*No Vote*	*Totals*
2002				
Changed Polling Place	856,097 41.684%	219,687 10.697%	977,991 47.619%	2,053,775 66.0%
Unchanged Polling Place	444,873 42.077%	107,580 10.175%	504,839 47.748%	1,057,292 34.0%
Total People	1,300,970 41.817%	327,267 10.519%	1,482,830 47.663%	3,111,067 100%
Difference	-0.393%	0.522%	-0.129%	
t-statistic	-6.65	14.21	-2.16	

TABLE 3: Estimated impact of Consolidation on Voter Turnout

2003 impact of Polling Place Change	Linear estimation, OLS (t-statistic)	Linear estimation, HLM (t-statistic, RSE)	Relative Change
Polling Place voting	-3.18% (-53.37)	-2.69% (-11.20)	4.77
Absentee Voting	+1.72% (40.73)	+1.40% (8.51)	4.79
Not Voting	+1.45% (24.82)	+1.28% (4.00)	6.21
2002 impact of Polling Place Change (i.e., Preexisting differences)			
Polling Place voting	-0.39% (-6.65)	-0.10% (-0.45)	14.8
Absentee Voting	+0.52% (14.21)	+0.40% (3.31)	4.29
Not Voting	-0.13% (-2.16)	-0.30% (-1.13)	1.91
2003-2002 impact of Polling Place Change			
Polling Place voting	-2.8% (-46.5)	-2.60% (-15.93)	
Absentee Voting	+1.2% (31.5)	+1.00% (10.61)	
Not Voting	+1.6% (27.2)	+1.61% (8.85)	

TABLE 4: Estimated impact of Consolidation on Voter Turnout, Conditioned on 2002 Voting Behavior

2003 impact of Polling Place Change, 2002 Polling Place Voters only	Linear estimation, OLS (t-statistic)	Linear estimation, HLM (t-statistic, RSE)	Relative Change
Polling Place voting	-4.5% (-56.28)	-4.22% (-21.44)	-10.14
Absentee Voting	+2.3% (41.57)	+1.87% (13.42)	4.50
Not Voting	+2.3% (33.62)	+2.36% (11.47)	5.67
2003 impact of Polling Place Change, 2002 Absentee Voters only			
Polling Place voting	-1.7% (-12.24)	-1.08% (-3.79)	-10.09
Absentee Voting	+1.4% (8.24)	+0.78% (2.21)	7.29
Not Voting	+0.3% (2.09)	+0.28% (1.01)	2.61
2003 impact of Polling Place Change, 2002 Non-Voters only			
Polling Place voting	-1.8% (-23.43)	-1.71% (-7.14)	-3.58
Absentee Voting	+0.7% (15.00)	+0.62% (5.35)	1.30
Not Voting	+1.1% (13.93)	+1.08% (3.64)	2.26

TABLE 5:
Selection Regression for Assignment of Registration Precincts

Variables	Basic Equation	t-statistic	With Interaction	t-statistic
Constant	0.583 (0.092)	6.353	0.553 (0.093)	5.945
Size of Precinct (People in 100's)	-0.012 (0.002)	-4.934	-0.012 (0.002)	-4.976
Absentee Voting (Fraction)	0.348 (0.086)	4.032	0.333 (0.087)	3.848
2002 Distance to Polling Place (thousandths of miles)	0.017 (0.003)	5.833	0.017 (0.003)	5.815
Age 60 or over (Fraction)	-0.171 (0.052)	-3.269	-0.042 (0.084)	-0.500
Age 40 or over (Fraction)	-0.029 (0.049)	-0.591	-0.034 (0.049)	-0.684
Polling Place Voting (Fraction)	-0.010 (0.044)	-0.227	0.056 (0.055)	1.016
Democrats (Fraction)	0.117 (0.096)	1.220	0.127 (0.096)	1.319
Republicans (Fraction)	0.091 (0.095)	0.950	0.104 (0.096)	1.089
Independents (Fraction)	0.081 (0.107)	0.749	0.089 (0.107)	0.829
Interaction: Fraction over Age 60 & Fraction Polling Place Voting	---	---	-0.00338 (0.00171)	-1.973
R^2	0.012		0.013	
N	8780		8780	

TABLE 6:
Outcome Estimates – Unadjusted, Change Scores, & Matching

	Analysis Method		
	Unadjusted Results	Change Score Results	Matching Results
Polling Place Voting	-3.09%	-2.75%	-3.05%
Absentee Voting	1.46%	1.11%	1.19%
Not Voting	1.62%	1.60%	1.88%
Number of People	2,722,240	2,722,240	2,722,240

TABLE 7: Hierarchical Multinomial Linear Models

	Category 1 – Polling Place Voting					
	Distance Grand-Centered			Distance Centered by Registration Precinct		
	Coefficient	Robust Standard Error	t-statistic	Coefficient	Robust Standard Error	t-statistic
Intercept	0.118	0.005	22.247	0.131	0.005	24.465
Changed Polling Place in 2003	**-0.122**	**0.009**	**-13.344**	**-0.181**	**0.007**	**-24.860**
Change in Distance 2002-2003	**-0.111**	**0.013**	**-8.471**	**-0.212**	**0.015**	**-13.698**
2002 Distance to Poll	0.050	0.012	4.198	-0.083	0.013	-6.165
2002 Polling Place Voter	2.244	0.005	472.334	2.254	0.005	482.555
2002 Absentee Voter	0.976	0.009	106.538	1.004	0.009	110.686
Major Party Registrant	0.263	0.004	62.169	0.273	0.004	64.326
* Permanent Absentee Voter	-1.033	0.017	-62.491	-1.034	0.017	-62.392
Age in 2003	0.105	0.001	168.084	0.102	0.001	163.729
Age^2/1000 in 2003	-0.949	0.006	-154.371	-0.926	0.006	-150.415

	Category 2 – Absentee Voting					
	Distance Grand-Centered			Distance Centered by Registration Precinct		
	Coefficient	Robust Standard Error	t-statistic	Coefficient	Robust Standard Error	t-statistic
Intercept	-1.322	0.008	-167.854	-1.358	0.008	-165.175
*** Change in Distance 2002-2003**	**0.173**	**0.023**	**7.545**	**-0.004**	**0.020**	**-0.212**
* 2002 Distance to Poll	0.312	0.028	11.216	0.083	0.020	4.150
2002 Polling Place Voter	1.544	0.008	205.648	1.546	0.007	219.087
2002 Absentee Voter	3.456	0.009	388.858	3.423	0.009	385.833
Major Party Registrant	0.282	0.007	40.020	0.304	0.007	44.125
Permanent Absentee Voter	1.093	0.012	87.896	1.115	0.012	89.994
Age in 2003	0.106	0.001	111.631	0.095	0.001	104.378
Age^2/1000 in 2003	-0.801	0.009	-88.973	-0.711	0.009	-83.535

FIGURE 1
Voter Decision Making

Net Benefits
of Voting at   0
Polls $(b-c^p)$

P

A

N

0

Net Benefits of Absentee Voting $(b-c^a)$

P: Polling Place Vote

**A: Absentee Vote**

N: No Vote
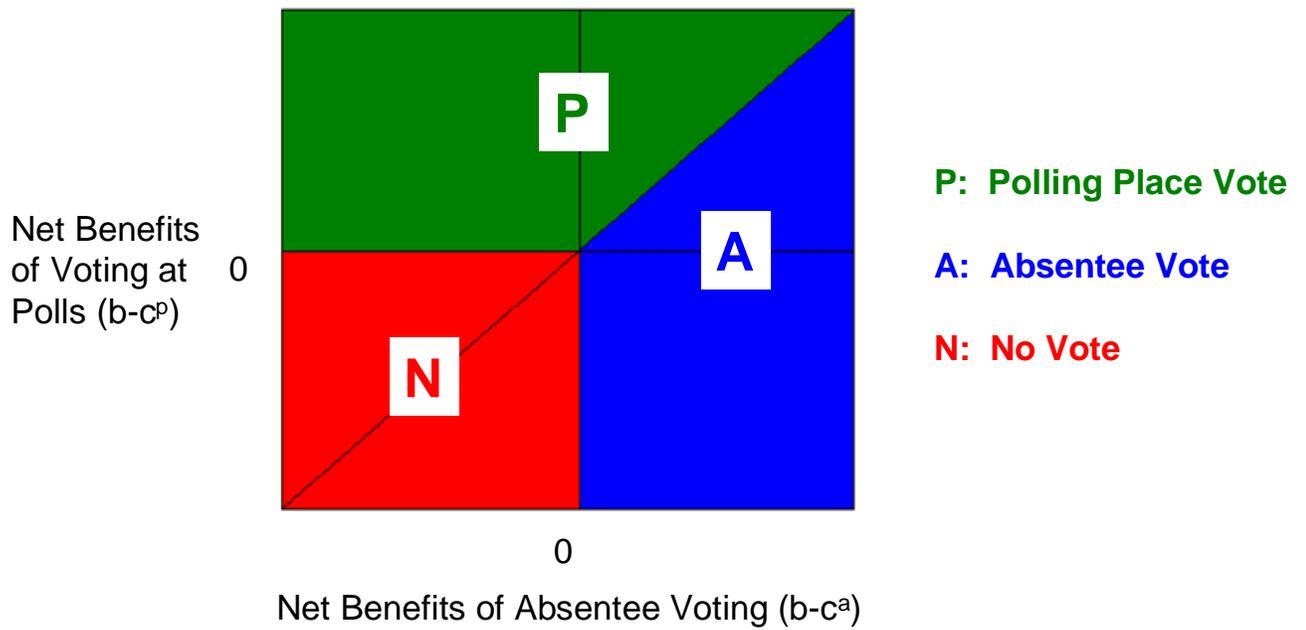
FIGURE 2
## Voter Decision Making With Shift in
Cost of Polling Place Voting

Net Benefits of Voting at Polls (b-c^p)

c^g

0

c^g{

0

Net Benefits of Absentee Voting (b-c^a)

P*

A$^#$

N$^#$

A

N

P*: Polling Place Vote

A: Absentee Vote

A$^#$: New Absentee Vote

N: No Vote

N$^#$: New No Vote

Figure 3

Distance to Polling Place in 2002

Distance in Hundredthsof Miles in 2002

Figure 4

Distance to Polling Place in 2003

Distance in Hundredths of Miles in 2003
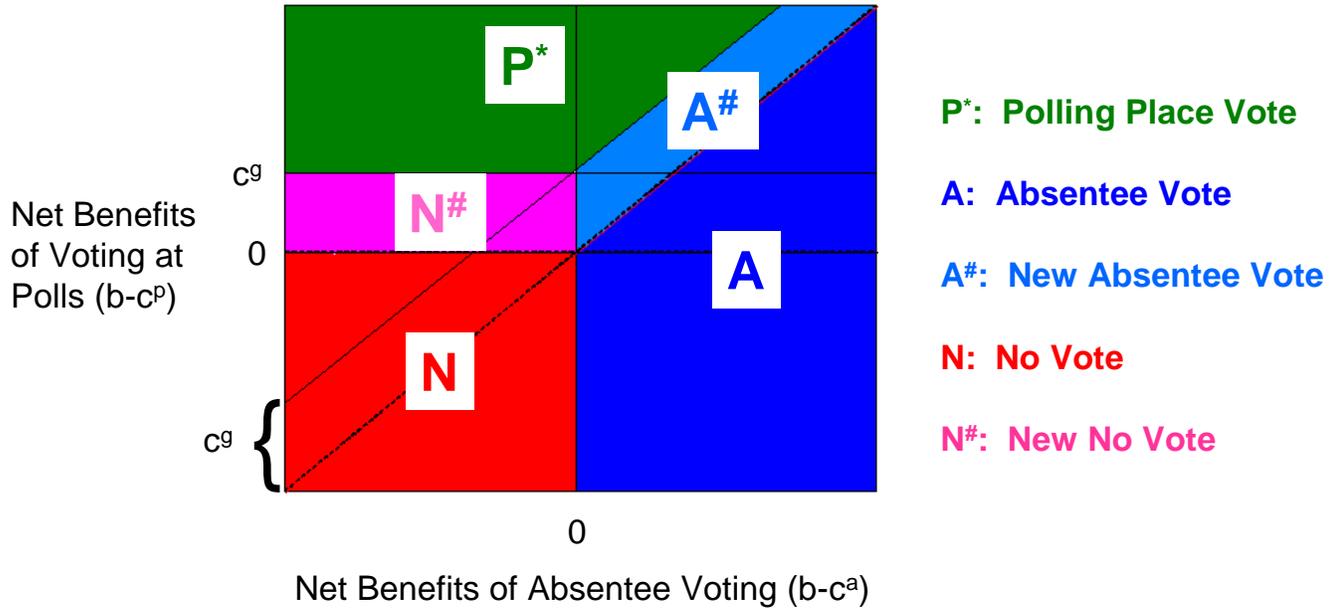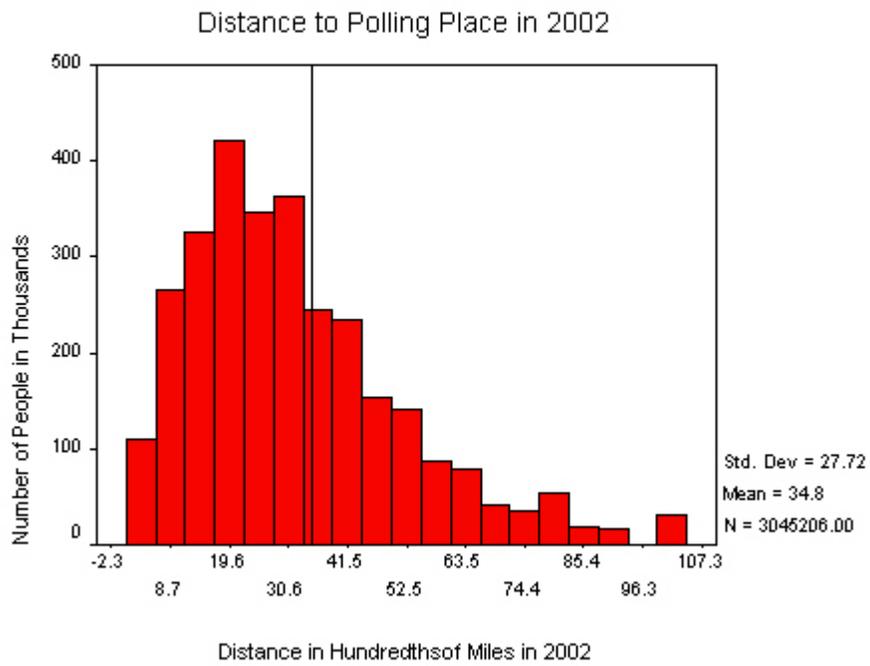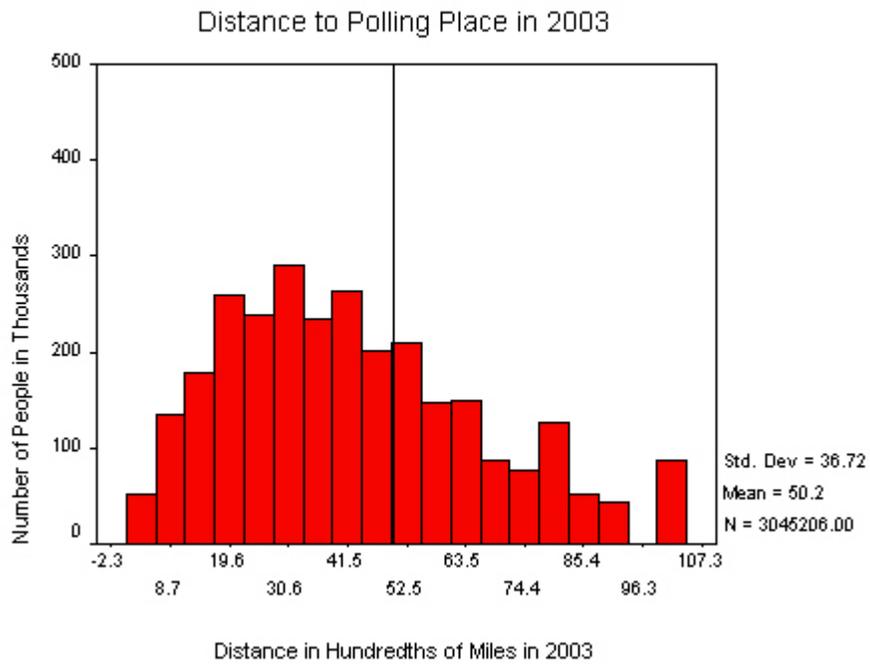
FIGURE 5:
2003 POLLING PLACE TURNOUT BY 2003 DISTANCE FROM POLL

Distance for Voter to 2003 Polling Place

Distance in hundredths of miles
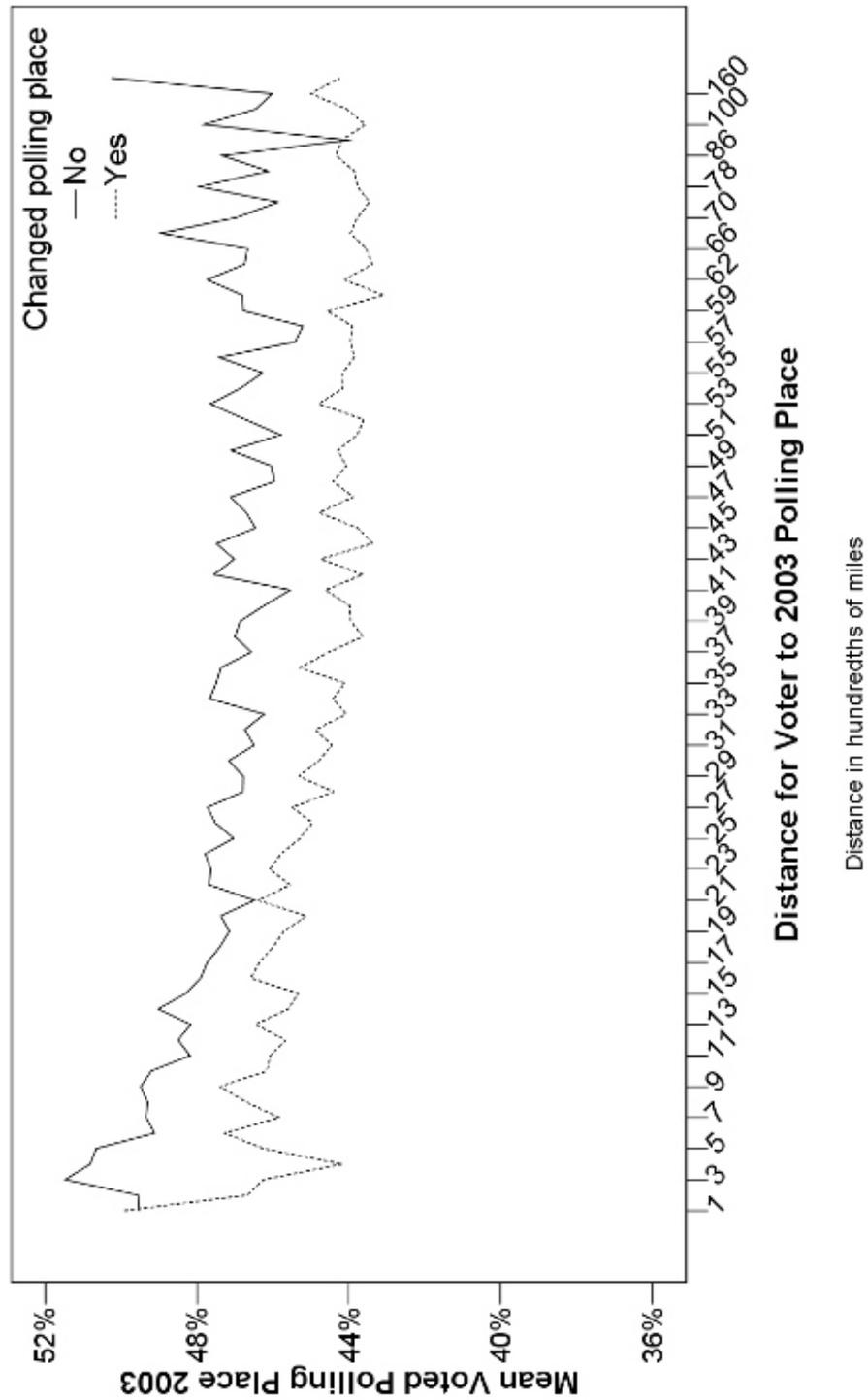
FIGURE 6:
2003 POLLING PLACE TURNOUT BY 2002 DISTANCE FROM POLL

Distance for Voter to 2002 Polling Place

Distance in hundredths of miles
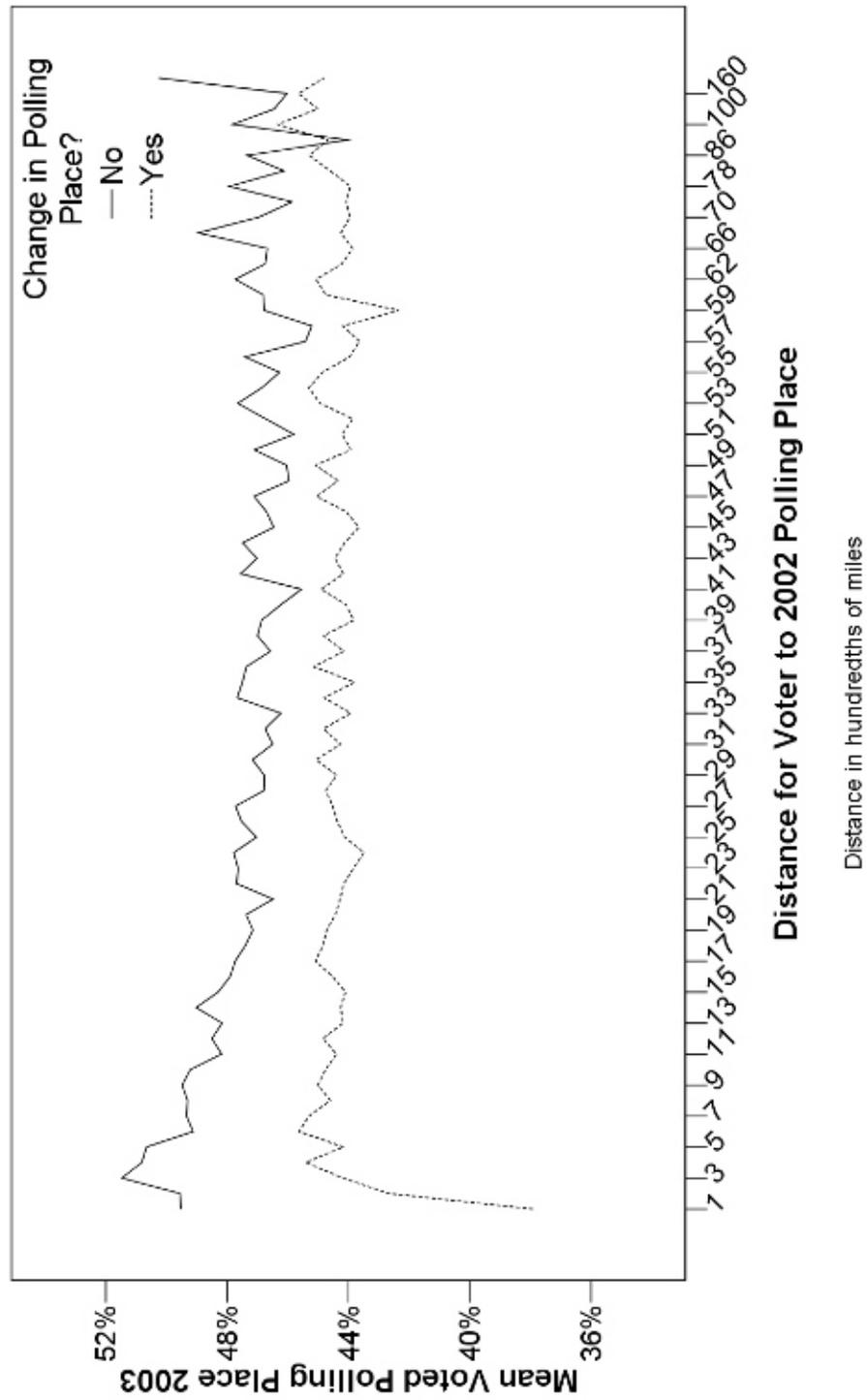
FIGURE 7:
2003 ABSENTEE VOTER TURNOUT BY 2002 DISTANCE FROM POLL
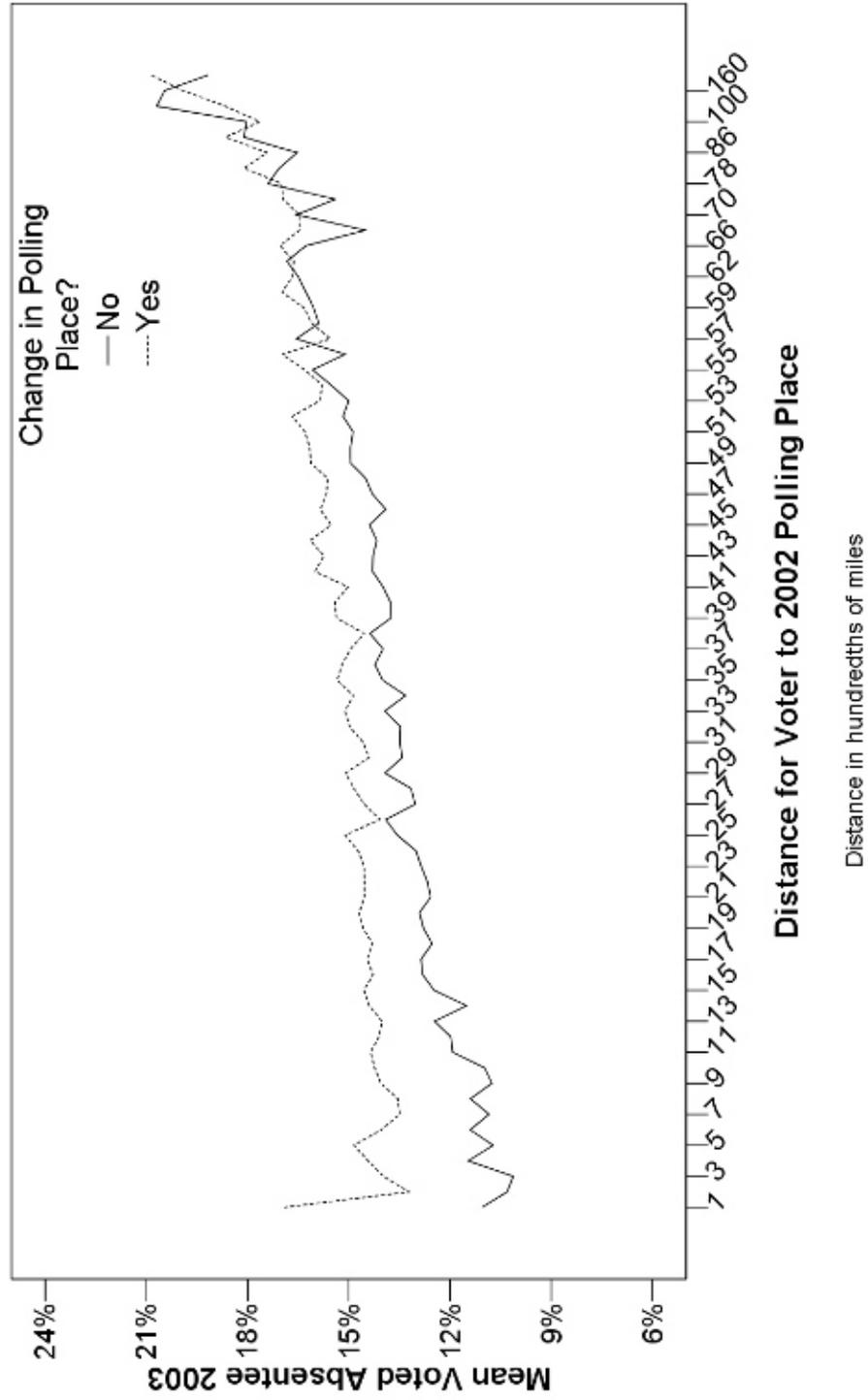
FIGURE 8:

CHANGE IN DISTANCE TO POLLING PLACE IN 2003 BY INITIAL DISTANCE IN 2002

Distance for Voter to 2002 Polling Place

Distance in hundredths of miles
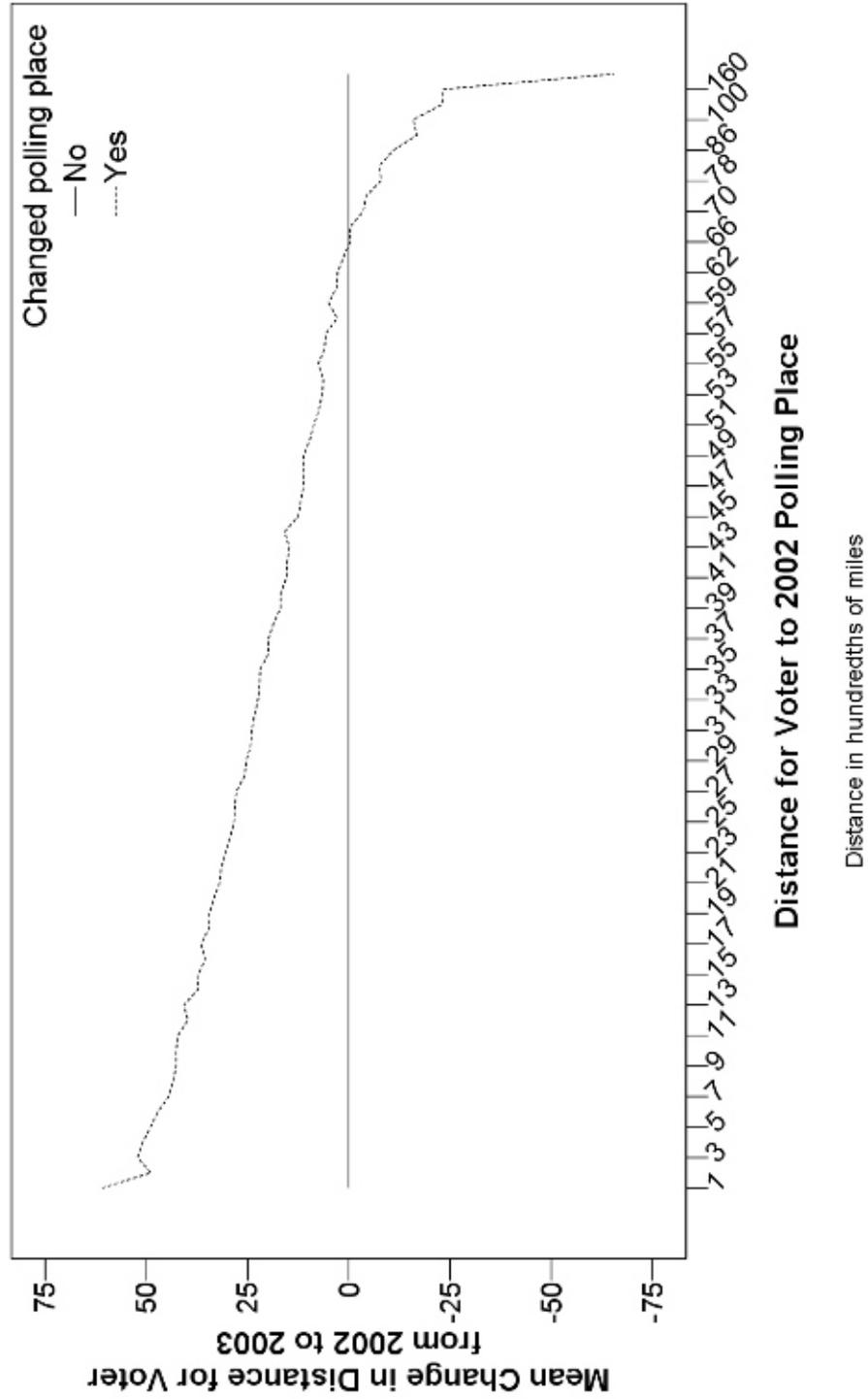
FIGURE 9:
2002 POLLING PLACE TURNOUT BY 2002 DISTANCE FROM POLL

Distance for Voter to 2002 Polling Place

Distance in hundredths of miles
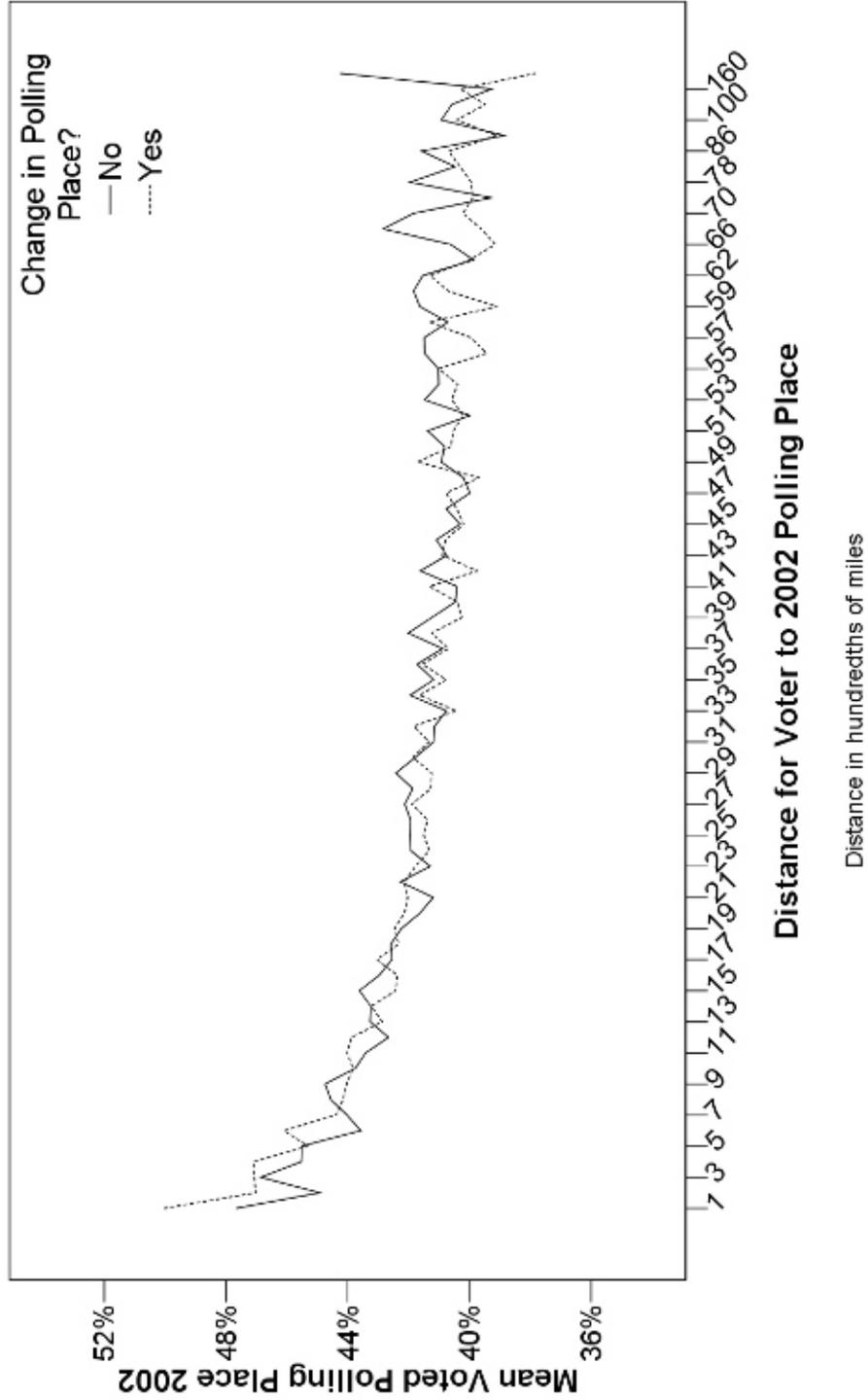
FIGURE 10:
2002 ABSENTEE VOTER TURNOUT BY 2002 DISTANCE FROM POLL

Change in Polling Place?
— No
---- Yes

Mean Voted Absentee 2002

Distance for Voter to 2002 Polling Place

Distance in hundredths of miles
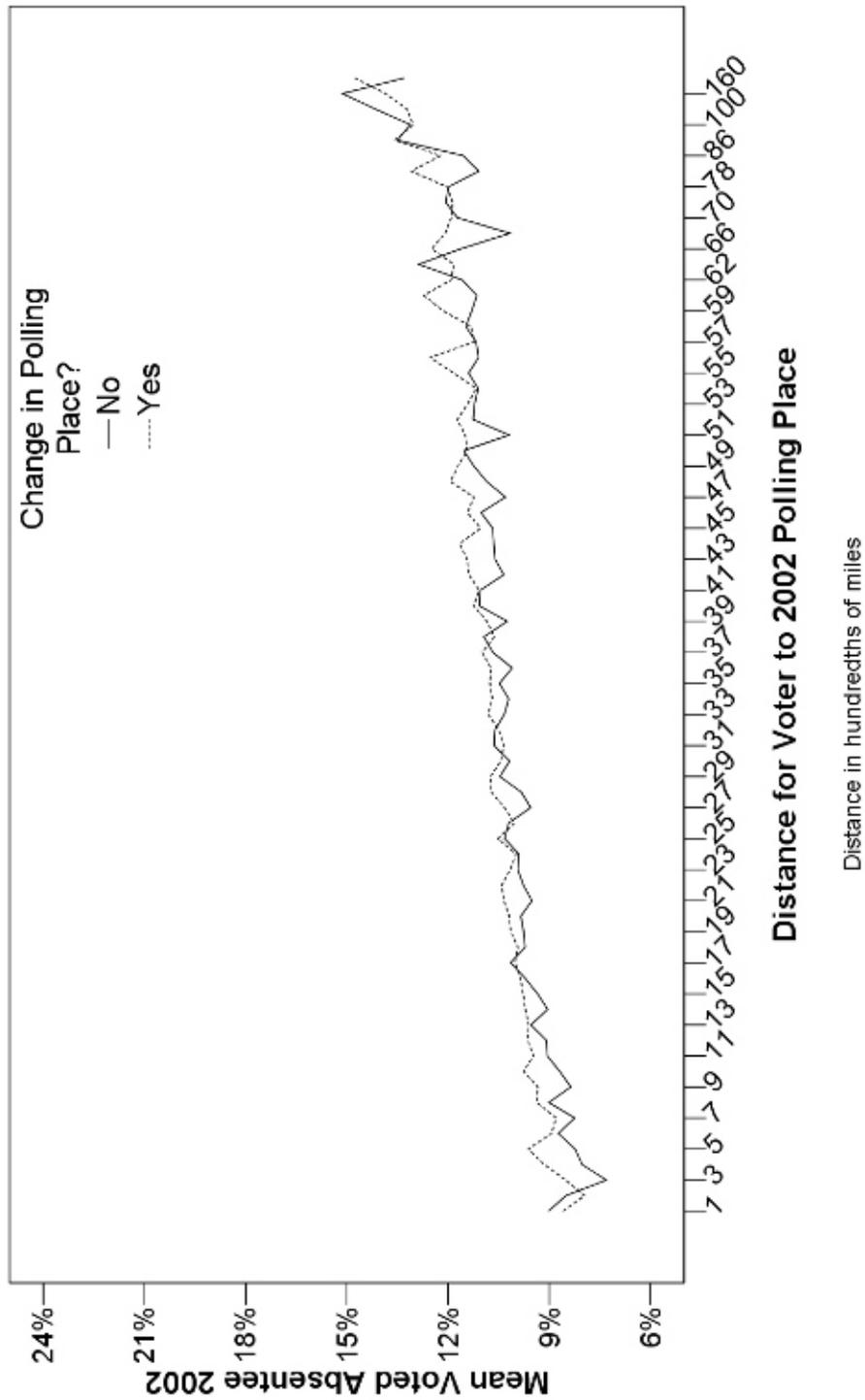
FIGURE 11: 2003 POLLING PLACE TURNOUT BY AGE
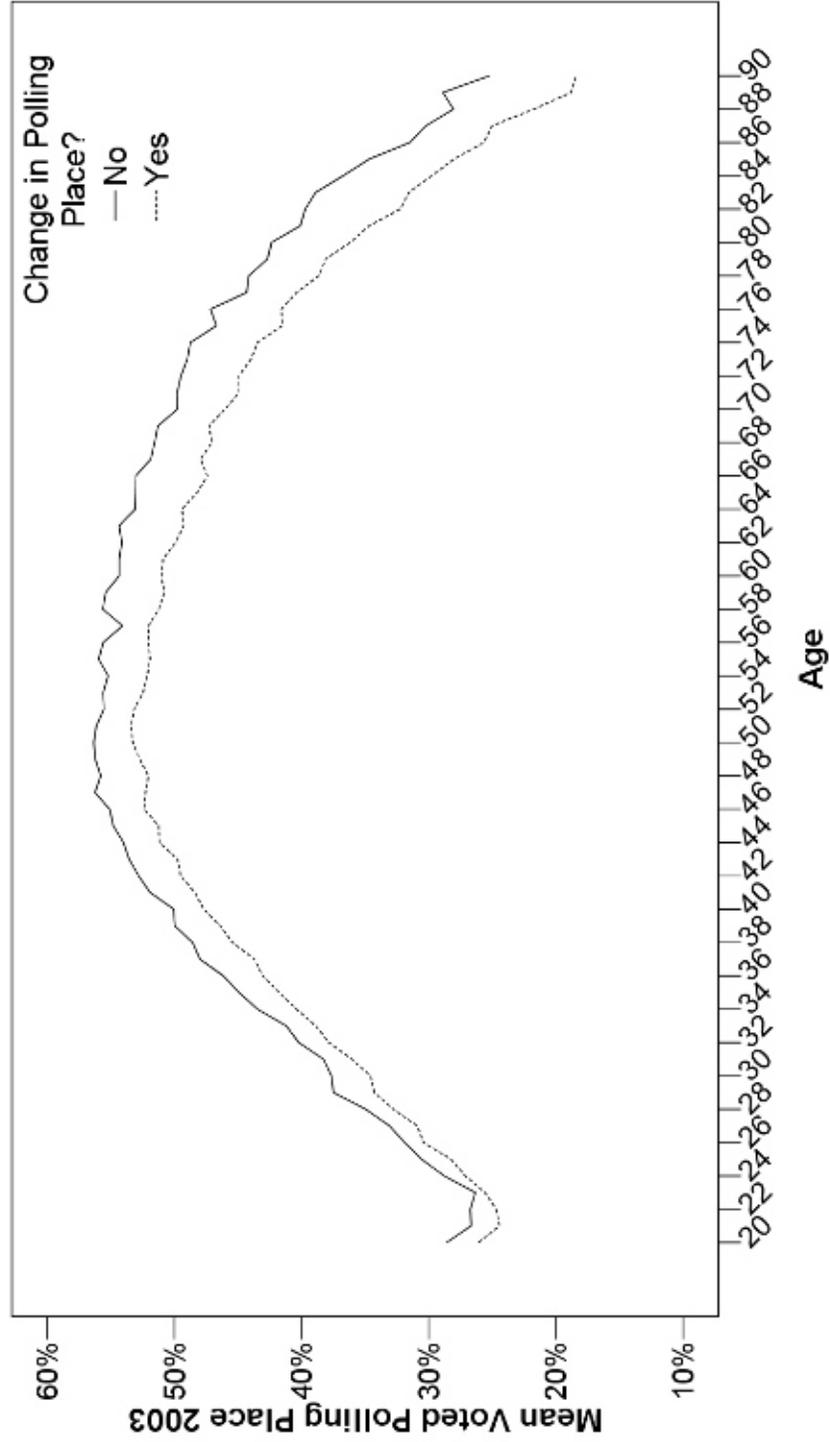
FIGURE 12: 2003 ABSENTEE VOTER TURNOUT BY AGE
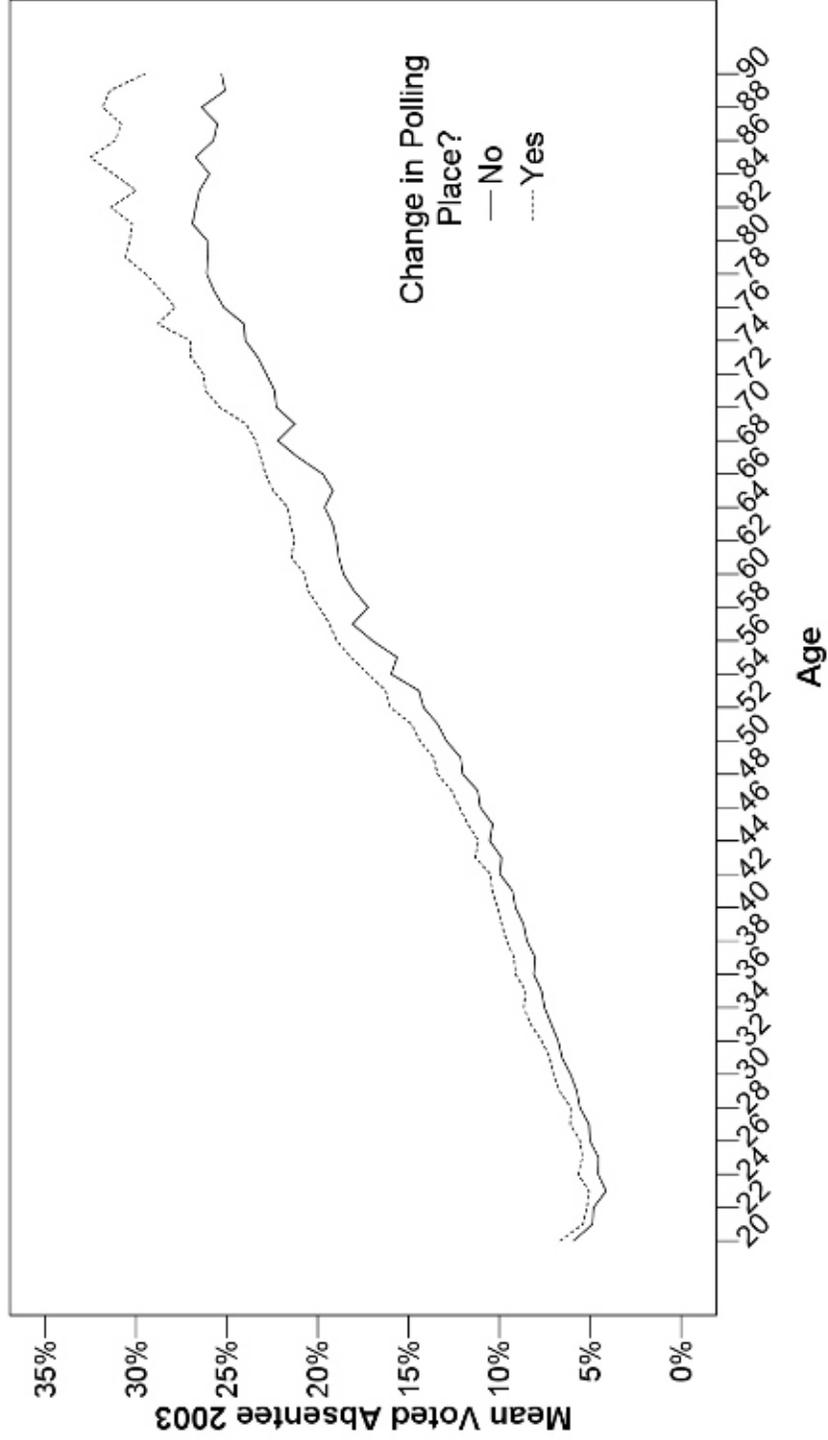
FIGURE 13: 2002 POLLING PLACE TURNOUT BY AGE
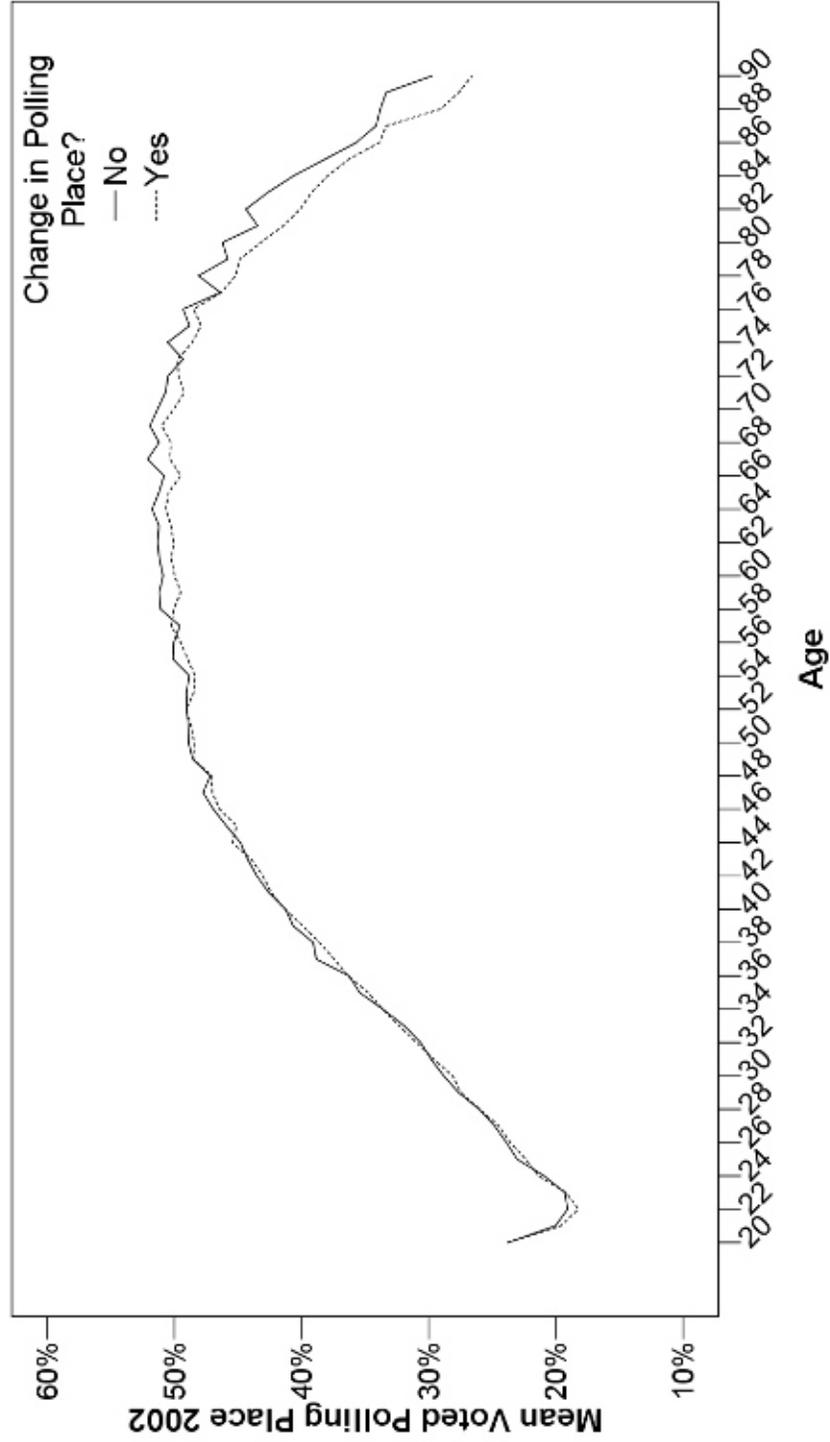
FIGURE 14: 2002 ABSENTEE VOTER TURNOUT BY AGE
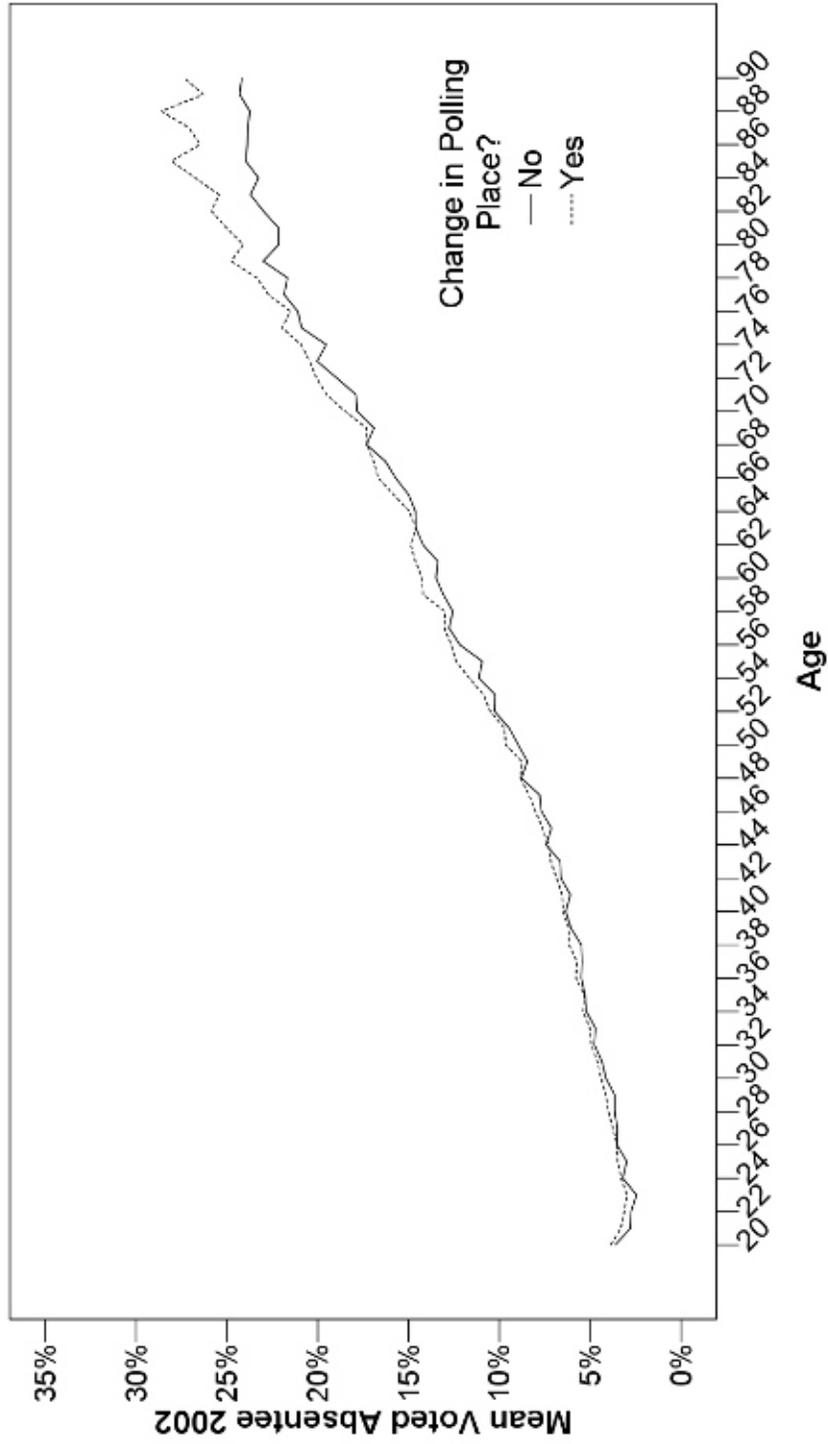
FIGURE 15: PERCENT CHANGED POLLING PLACE
BY 2002 DISTANCE TO POLL

Mean Changed polling place

Distance for Voter to 2002 Polling Place

Distance in hundredths of miles
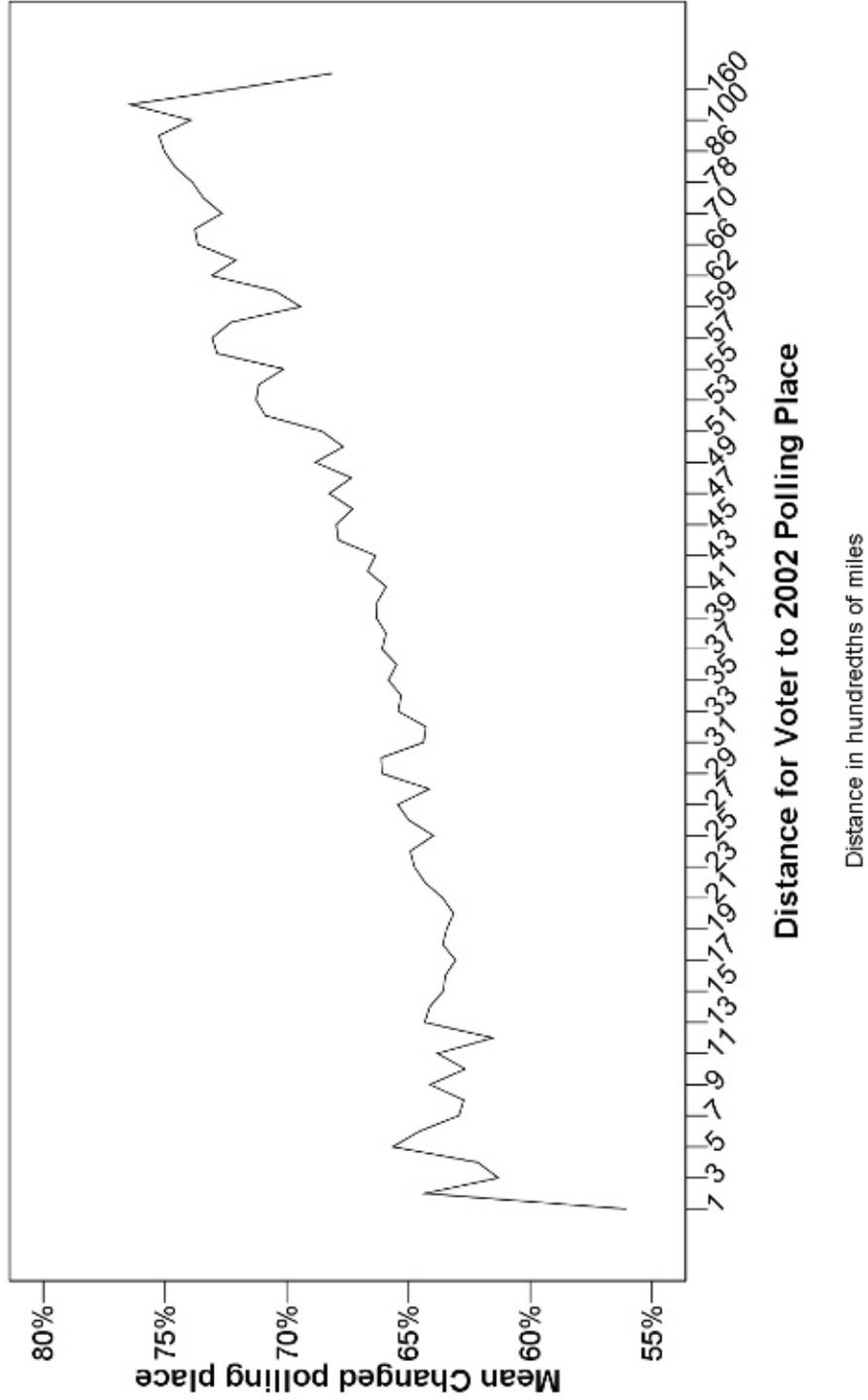
FIGURE 16: PERCENT CHANGED POLLING PLACE BY AGE
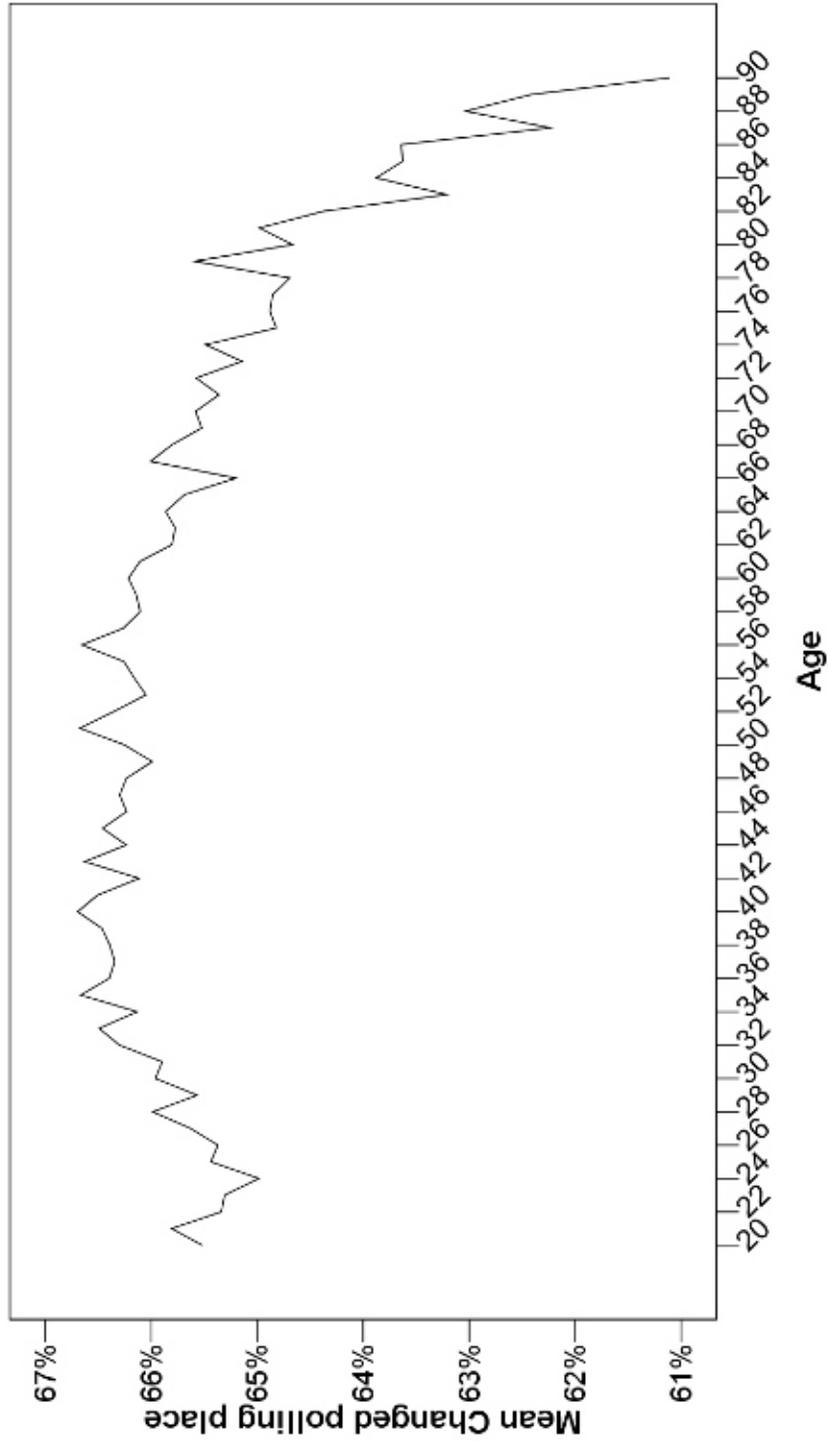
Figure 17

Number of Registration Precincts by

% of People Changing Polling Place


Percentage of People who Changed Polling PLace
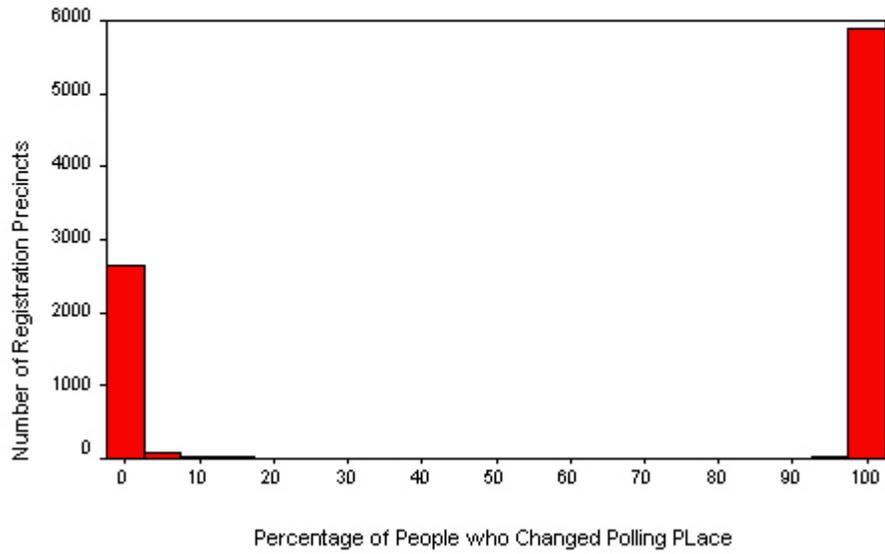
Figure 18

Blowup of Number of Registration Precincts

by % of People Changing Polling Place

Percentage of People Who Changed Polling Place