Using Cluster Randomized Field Experiments to Study Voting Behavior

By
KEVIN ARCENEAUX

Voter mobilization experiments are often conducted using individual-level randomization, which can be difficult to implement. A simpler approach is to randomly assign voting precincts, rather than individuals nested within them, to treatment and control groups. Not only is it easier and potentially less expensive to implement, it may allow researchers to study vote preference effects without collecting survey data. This article explores various methodological concerns that researchers should consider when designing and analyzing precinct-level experiments. These concerns are illustrated using data from a precinct-level randomized field experiment conducted in Kansas City, Missouri.

Keywords: voter mobilization; vote choice; clustering; cluster randomization; group randomization; field experiments

Voter mobilization experiments typically rely on individual-level randomization. Lists of registered voters are obtained and individuals are randomly assigned to treatment groups (e.g., Eldersveld 1956; Gerber and Green 2000; Miller, Bositis, and Baer 1981). The advantage of individual-level randomization is that the design generates a great deal of statistical power, but the disadvantage is that it may be accompanied by high costs. For door-to-door canvassing, individual-level randomization requires a great deal of care in constructing walk lists that avoid contact with members from other treatment groups or the control group and necessitates training canvassers to avoid contact with individuals who are not specifically on their target lists.

Kevin Arceneaux (Ph.D., Rice University, 2003) is currently a postdoctoral fellow in the Yale University Institution for Social and Policy Studies. His research focuses on questions of democratic representation, paying special attention to the factors that influence democratic accountability. To date, his research has been directed at exploring elite-mass linkages and electoral behavior. He is currently studying the role campaigns play in shaping attitudes and voting behavior, and he hopes to extend this research to include the news media in the near future.

DOI: 10.1177/0002716205277804

It also may increase the geographic area to be covered since canvassers will be skipping houses that have been assigned to the control group. Consequently, individual-level randomization may be difficult to execute on limited budgets or with volunteer canvassing pools. Moreover, this design may be a difficult sell to campaigns that do not want to restrict incidental contact or that wish to keep walk lists geographically compact.

An alternative is for researchers to use cluster randomization. In this design, clusters, rather than the individuals nested within them, are randomly assigned to treatment conditions. For voter mobilization experiments, this sort of randomization can be easily executed by assigning voting jurisdictions, such as precincts, to treatment groups. This design is far easier to implement since canvassers need not be assigned to specific houses or individuals.

Another benefit to randomly assigning clusters is that researchers can study the effect of voter mobilization on voter preferences, as well as turnout. Both individual and cluster randomized experiments can use publicly available voter files to collect data on whether subjects voted and gauge the impact of voter mobilization campaigns on the decision to vote. Yet because information on individual voting decisions is not publicly available, individual randomized experiments are unable to assess the impact of mobilization campaigns on voter preferences without independently collecting these data. Postelection surveys are adequate but can be quite expensive. Furthermore, the effect size may be distorted by measurement error because it is not clear how to allocate respondents who refuse to reveal their vote choices. In contrast, experiments randomized at the level of the voting precinct can use publicly available election reports to study preference effects. Not only is this an inexpensive technique, it does not confront the degree of measurement error that surveys do.

The benefits of group randomized experiments come at the cost of statistical power. Because voting decisions are likely to be correlated within precincts, the number of independent observations, or the "effective *N*," is lower than it would be in an identical individual randomized experiment (Stoker and Bowers 2002). Consequently, there is greater uncertainty associated with estimates in cluster randomized designs. In the pages that follow, I will describe methods to account for this uncertainty and reduce it, which will be illustrated with the help of a voter mobilization experiment randomized at the precinct level.

Statistical Considerations for Field Experiments

Because subjects are randomly assigned to treatment and control groups, the treatment effect for mobilization experiments can be estimated by comparing the voting rates of the two groups. Randomization ensures that individuals in the treatment group have roughly the same baseline probability of voting as those in the control group, so a difference in voting rates can be attributed to the campaign effort. Because campaigns rarely, if ever, contact 100 percent of the treatment group, a simple comparison of voting rates between the two groups estimates the intent-to-treat (ITT) effect. One can estimate the average treatment-on-treated (ATT) effect by dividing the ITT effect by the contact rate, which is identical to a two-stage least squares (2SLS) analysis in which individual turnout is regressed on whether the individual was contacted and random assignment is the instrumental variable (Gerber and Green 2000).[1]

Individual-level randomization may be difficult to execute on limited budgets or with volunteer canvassing pools.

These approaches to estimating the ITT and ATT can also be used when clusters (e.g., precincts) are the unit of randomization, but not without adjusting the standard errors. The regression models that are appropriate for individual random assignment assume that observations are independent from one another. With cluster random assignment, observations within clusters are likely not independent. Citizens who live in the same precinct likely share similar predilections to vote. If this correlation is not accounted for, standard errors will be biased downward. One can account for intracluster correlation by using a robust variance estimator (Donner and Klar 2000), which is available in many software packages. The robust variance estimator allows observations to be correlated within clusters when calculating the variance-covariance matrix (White 1980).[2]

The consequence of intracluster correlation is a reduction in the effective *N* and a loss of statistical power. Researchers can overcome some of the loss of statistical power by including covariates that are predictive of the dependent variable (Murray 1998; Raudenbush 1997). In the case of voter turnout, past voting behavior, age, gender, and party registration are often available in the voter file; these covariates, which are highly predictive of turnout, increase the precision of the treatment effect estimate.

In addition to adding covariates, researchers can increase the power of cluster randomized experiments by increasing the number of clusters included in the design. The increase in effective *N* obtained from additional clusters depends on how highly correlated individuals are within clusters. Researchers should conduct a power analysis prior to random assignment to calculate the optimal number of clusters to be assigned. Bloom, Bos, and Lee (1999) and Raudenbush (1997) provided excellent illustrations of power analysis for cluster random assignment.

An Illustration

The study

To illustrate cluster random assignment in practice, I analyze data from a voter mobilization experiment conducted in Kansas City, Missouri. A budget crisis forced the Kansas City Transit Authority to place a proposal on the November 2003 ballot requesting a 3/8 percent increase in the city's sales tax to increase revenue. If the ballot measure were not approved, bus service would have to be curtailed. The Association for Community Organizations for Reform Now (ACORN), a community group that advocates on behalf of low-income families, organized a voter mobilization campaign in support of the tax that featured door-to-door canvassing.

The increase in the total number of voters may work to lower the percentage of support for the ballot proposition even if more yeses are mobilized than nos.

ACORN selected twenty-eight primarily African American precincts to be included in the study. Half (fourteen) were randomly assigned to be canvassed, while the other half were assigned to the control group. Relying on a pool of seventy-five predominately African American canvassers between eighteen and twenty-five years old, ACORN began a month-long intensive campaign to educate voters about the proposed sales tax proposal and encourage them to vote. Canvassers walked the treatment precincts twice. The first walk through the precincts focused on education. Door-knock scripts were aimed at persuading voters to vote for the proposed sales tax increase rather than face serious cuts in bus service. Many individuals in the precincts studied (both treatment and control) relied heavily on buses for transportation, and 84.2 percent of those contacted in the first walk expressed support for the proposal. The second walk was conducted a week out from the election and focused on voter mobilization. Contacts from the first walk who expressed support for the tax increase were reminded about the sales tax proposal and encouraged to vote.

TABLE 1
BALANCE OF PAST VOTING BEHAVIOR IN TREATMENT AND CONTROL GROUPS

Past Election	Treatment	Control	Difference	*t*-Statistic
2003 primary	11.9	13.0	–1.1	–0.74
2003 municipal	34.6	36.9	–2.3	–0.77
2002 general	63.1	65.2	–2.1	–0.91
2002 primary	22.2	21.6	0.6	0.31
2002 municipal	14.8	15.6	–0.8	–0.45
2001 special	26.3	26.9	–0.6	–0.23
2000 general	85.8	86.8	–1.0	–0.42
2000 primary	21.1	20.0	1.1	0.52
2000 municipal	13.8	14.4	–0.6	–0.36
2000 special	4.4	0.4	4.0	2.15
Number of observations	14	14		

Randomization check

It is important to ascertain whether random assignment was conducted properly. There should be little to no differences between treatment and control groups. Because it is a simple matter to reassign groups, it is optimal to conduct this analysis before the experiment begins. The randomization check should be performed *at the level of randomization*, rather than the level of analysis.[3] Thus, I check randomization at the precinct level. Table 1 displays precinct-level voting rates in previous elections for the treatment and control groups. There are no statistically significant differences between the treatment and control group precinct-level voting rates in the past nine most recent elections. Indeed, a joint test of significance, which tests whether past voting behavior taken together affects random assignment, shows that previous voting behavior is not related to random assignment (p = .49).[4]

Estimating voter turnout effects

Table 2 displays the ITT and ATT effects for various models. All estimates were obtained using a linear probability model (i.e., ordinary least squares [OLS] and 2SLS). Nonlinear probability models produce similar results, and are not reported here to conserve space. The standard errors in columns 1 and 4 were estimated using the common assumptions of homogeneity and independence in errors. Robust variance estimators are used in the remaining columns. Columns 3 and 6 add covariates to the model.

Notice that standard errors in columns 1 and 4 are drastically underestimated, compared to the standard errors that correct for intracluster correlation in columns 2 and 5. The robust estimates of the standard errors are nearly three times

TABLE 2

THE IMPACT OF VOTER MOBILIZATION ON TURNOUT (DEPENDENT VARIABLE = TURNOUT)

	Intent-to-Treat (ITT) Effects			Average Treatment-on-Treated (ATT) Effects		
Variable	1	2	3	4	5	6
Treatment	4.4	4.4	5.3	7.0	7.0	8.5
	(0.9)	(2.5)	(1.7)	(1.5)	(3.9)	(2.7)
2003 primary			21.7			21.6
			(2.3)			(2.3)
2003 municipal			24.7			24.4
			(1.2)			(1.2)
2002 general			12.4			11.9
			(0.9)			(0.8)
2002 primary			8.7			8.8
			(1.7)			(1.7)
2002 municipal			7.9			8.0
			(1.8)			(1.7)
2001 special			6.0			6.0
			(1.4)			(1.4)
2000 general			4.6			4.0
			(1.3)			(1.3)
2000 primary			6.1			6.0
			(1.8)			(1.8)
2000 municipal			3.2			3.3
			(1.5)			(1.5)
2000 special			6.1			5.3
			(2.3)			(2.3)
Constant	29.1	29.1	–1.6	29.1	29.1	–0.6
	(0.7)	(0.2)	(1.8)	(1.4)	(2.1)	(1.6)
Number of observations	9,712	9,712	9,712	9,712	9,712	9,712
F	21.89	3.06	626.97	22.16	3.16	622.00
R^2	.002	.002	.32	.01	.01	.32

NOTE: Standard errors reported in parentheses. Columns 1 and 4 do not correct the standard errors, while the rest do.

larger than the uncorrected standard errors. This stark difference illustrates the need to correct for intracluster covariance and provides a concrete example of Cornfield's (1978, 101) warning, "Randomization by cluster accompanied by an analysis appropriate to randomization by individual is an exercise in self-deception."

Columns 3 and 6 demonstrate that the addition of predictive covariates reduces the corrected standard errors considerably by shrinking the amount of overall unexplained variance. These standard errors are 30 percent lower than the without-covariates model. The inclusion of covariates also happens to increase the size

TABLE 3
COMPARING INDIVIDUAL-LEVEL AND PRECINCT-LEVEL ANALYSES

	Individual Level	Precinct Level
Intent-to-treat (ITT) effects		
Without covariates	4.4 (2.5)	3.6 (2.4)
With covariates	5.3 (1.7)	5.1 (1.6)
Average treatment-on-treated (ATT) effects		
Without covariates	7.0 (3.9)	6.5 (4.1)
With covariates	8.5 (2.7)	9.4 (3.0)

NOTE: Standard errors reported in parentheses. Cluster corrected standard errors are reported for the individual-level models.

of the effects. Asymptotically, though, these estimates would be unchanged with the inclusion of covariates because the without-covariates estimate is consistent but inefficient. Given the small number of clusters (twenty-eight), there is some movement in the with-covariates estimate, but it falls within the error range of the without-covariates estimate. Nevertheless, as this example demonstrates, covariates can be a powerful tool to increase statistical power in group randomized designs.

Aggregate analysis

Because precincts are the unit of randomization, it is also appropriate to conduct the analysis at this level. By treating precincts as the until of analysis, one does not have to account for intracluster correlation. However, this approach also severely reduces the effective *N*, resulting in less precise estimates. To illustrate, Table 3 compares the individual-level results reported in Table 2 to corresponding analyses conducted at the precinct level. Note that the standard errors are approximately the same for both analyses. The benefits of individual-level data are most noticeable in the 2SLS analysis. Here, the individual-level model correctly calculates the contract rate for the sample, while the precinct-level analysis relies on cluster averages of the contact rate. Consequently, individual-level data obtained from cluster randomized experiments, if available, should be used.

Estimating voter preference effects

Because electoral outcomes are reported by precinct, I can also assess whether ACORN successfully convinced citizens to vote for the sales tax proposal. This analysis can be done by calculating the marginal difference between yes and no precinct-level votes for the treatment and control groups and comparing them. The marginal vote differential (MVD) is calculated as follows:

$$MVD_j = \frac{Y_j - N_j}{R_j} \times 100, \tag{1}$$

where Y_j is the number of yes votes in the jth precinct, N_j is the number of no votes in the jth precinct, and R_j is the total number of registered voters in the jth precinct. One should use this statistic rather than the percentage of support in the precinct, because a campaign may mobilize both yes and no voters. The increase in the total number of voters may work to lower the percentage of support for the ballot proposition *even if more yeses are mobilized than nos*.

To illustrate this possibility, consider a hypothetical experiment in which the treatment and control groups each consist of 100 individuals who will vote and 200 who will not. Because of random assignment, baseline support for an issue is the same in both groups: 70 voters will vote for the issue and 30 will vote against it. A campaign advocating this issue contacts individuals in the treatment group, mobilizing 60 of the 200 voters who had planned on not voting. They are successful in persuading 40 of these new voters to support their issue (assume there are no changes in baseline voter preferences). Despite their success in mobilizing more new supporters than opponents, a naive inspection of the percentage of support would suggest failure: 70 percent of the control group supports the issue ([70/100] × 100), while only 68 percent support it in the treatment group ({[70 + 40]/[100 + 60]} × 100). In contrast, the MVD provides a more accurate picture. In the control group it is 13 percent ({[70 – 30]/300} × 100) and in the treatment group it is 20 percent ({[(70 + 40) – (30 + 20)]/300} × 100).

As with the turnout analysis, I add covariates to boost the power of the analysis. Since the number of precincts in the study is small, I included data from 154 precincts in Kansas City for which vote outcomes were available across all elections, which allows more degrees of freedom to support the addition of covariates. The inclusion of these additional precincts will not affect the interpretation of the treatment effect because their inclusion will be modeled. They are only included to help the covariates improve the fit of the model, reducing unexplained variance and increasing power. I use the following OLS model to fit these data:

$$MVD_j = \beta_0 + \beta_1 T_j + \beta_2 J_j + \beta_3 S_j + \sum_k b_k MVD_{jt-1} + e_j, \qquad (2)$$

where T_j indicates whether the jth precinct is in the treatment or control group {1, 0}, J_j indicates whether the jth precinct was combined with another one outside of the study {1, 0} (the four precincts combined across treatment and control group are removed), S_j indicates whether the jth precinct is in the study {0, 1}, and MVD_{jt-1} are covariates for past precinct voting behavior. These results are displayed in Table 4. S_j effectively models the inclusion of precincts outside of the study so that the interpretation of the treatment effect is the same even if they had not been included. The MVD ITT effect of 0.9 percent suggests that the ACORN campaign successfully turned out supporters for the proposal, but the standard error for this coefficient is too large to rule out chance.

Combining estimates from the turnout ITT effect (see Table 2) and the MVD ITT effect (Table 4) allows researchers to estimate the extent to which a campaign

TABLE 4
THE IMPACT OF VOTER MOBILIZATION ON VOTE PREFERENCE (DEPENDENT VARIABLE = MARGINAL VOTE DIFFERENTIAL)

Variables	Intent-to-Treat (ITT) Effect
Treatment	0.9 (0.9)
Combined precinct	–3.4 (1.2)
In study	3.0 (0.7)
MVD drug tax (2003)	7.1 (12.7)
MVD Senate (2002)	2.4 (1.9)
MVD cigarette tax (2002)	–13.6 (5.2)
MVD revitalization tax (2002)	–12.6 (3.7)
MVD convention tax (2002)	–1.6 (7.8)
MVD public bond (2002)	17.6 (8.0)
MVD fuel tax (2002)	(41.9) (8.8)
Constant	–1.2 (0.7)
Number of observations	154
F	86.45
R^2	.85

NOTE: Standard errors reported in parentheses. MVD = marginal vote differential.

mobilizes supporters. Let α = the proportion of supporters among those mobilized by the campaign, where α ranges from 0 to 1 (0 = all new voters who voted against the campaign and 1 = all new voters who voted for the campaign). Let t = turnout ITT effect. With these definitions, the MVD for the treatment group can be expressed as follows:

$$MVD_{treatment} = \frac{(Y + \alpha t r_T) - [N + (1 - \alpha) t r_T]}{R_T} \times 100, \tag{3}$$

where N = number of no votes in the absence of a campaign intervention, Y = number of yes votes in the absence of a campaign intervention, R_T = number of registered voters in the treatment group, and r_T = number of targeted registered voters in the treatment group. The MVD for the control group reflects baseline support,

$$MVD_{control} = \frac{Y - N}{R_C} \times 100, \tag{4}$$

where R_C = number of registered voters in the control group. The MVD ITT effect is simply the difference in MVD between the treatment and control group,

$$MVD_{ITT} = MVD_{treatment} - MVD_{control}. \tag{5}$$

Simplifying this equation produces

$$MVD_{ITT} = \frac{N\left(R_T - R_C\right) - R_T Y + R_C\left(2\alpha t r_T - t r_T + Y\right)}{R_C R_T}. \tag{6}$$

As a result, we can use estimates of t and MVD_{ITT} to estimate α. Solving for α in equation (6),

$$\alpha = \frac{N\left(R_C - R_T\right) + R_T Y + R_C\left(t r_T - Y + R_T MVD_{ITT}\right)}{2 R_C t r_T}. \tag{7}$$

The turnout ITT effect reported in Table 2 (with covariates) is 5.3, and the MVD_{ITT} effect reported in Table 4 is 0.9, N = 185, Y = 1,706, R_T = 15,839, R_C = 13,290, and r_T = 4,933.[5] Substituting these quantities into equation (7), α = 0.78 (see equation [8]), suggesting that 78 percent of the new voters mobilized by ACORN supported the ballot proposition. In concrete terms, this analysis shows that ACORN mobilized 261 people to vote (4,933 targeted individuals times 5.3 percent ITT turnout effect). Consequently, the campaign *netted* 146 yes votes ([0.78 – (1 – 0.78)] × 261).

$$\frac{185\left(13,290 - 15,839\right) + 15,839 * 1,706 + 13,209\left(5.3 * 4,933 - 1,706 + 15,839 * 0.9\right)}{2 * 13,209 * 5.3 * 4,933} = 0.78. \tag{8}$$

Conclusion

This article illustrates how researchers can use precinct-level randomization to conduct vote mobilization experiments. The design offers a number of benefits, including ease of implementation and the ability to readily analyze the impact of campaigns on vote preference as well as turnout. Yet researchers should be aware of special statistical considerations. Care should be taken when designing the experiments to maximize the effective number of observations and ensure greater statistical power. Researchers should also collect data on covariates that are predictive of outcomes, again boosting statistical power. Finally, the statistical analysis should correct for the clustered nature of the data when calculating standard errors. Those who fail to correct standard errors for intracluster correlation risk making conclusions their data do not support.

Notes

1. Two-stage least squares assumes a linear probability model. Nonlinear probability models can also be used to estimate the average treatment-on-treated (ATT) (Wooldridge 2002).

2. Bingenheimer and Raudenbush (2004, 54-60) and Donner and Klar (2000, 116-17) recommended using a multilevel model to estimate treatment effects. However, these models can only be used to estimate the intent-to-treat (ITT) effect and assume that the intracluster correlation is constant across clusters. Robust variance estimators can be used in models that estimate the ATT effect and allow the intracluster correlation to vary across clusters.

3. Since observations are not independent within precincts, a randomization check at the individual level, which is the level of analysis in this example, would be more likely to indicate that the randomization had failed even if it had not.

4. A joint test of significance can be conducted by regressing random assignment on available covariates (e.g., past voting behavior) and inspecting the statistical significance of the overall fit statistic.

5. N and Y were obtained from the control group.

References

Bingenheimer, Jeffrey B., and Stephen W. Raudenbush. 2004. Statistical and substantive inferences in public health: Issues in the application of multilevel models. *Annual Review of Public Health* 25:53-77.

Bloom, Howard S., Johannes M. Bos, and Suk-Won Lee. 1999. Using cluster random assignment to measure program impacts: Statistical implications for the evaluation of education programs. *Evaluation Review* 23 (4): 445-69.

Cornfield, J. 1978. Randomization by group: A formal analysis. *American Journal of Epidemiology* 108 (2): 100-102.

Donner, Allan, and Neil Klar. 2000. *Design and analysis of cluster randomization trials in health research*. London: Arnold.

Eldersveld, Samuel J. 1956. Experimental propaganda techniques and voting behavior. *American Political Science Review* 50 (1): 154-65.

Gerber, Alan S., and Donald P. Green. 2000. The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review* 94 (3): 653-64.

Miller, Roy E., David A. Bositis, and Denise L. Baer. 1981. Stimulating voter turnout in a primary: Field experiment with a precinct committeeman. *International Political Science Review* 2 (4): 445-60.

Murray, David M. 1998. *Design and analysis of group-randomized trials*. Oxford: Oxford University Press.

Raudenbush, Stephen W. 1997. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* 2 (2): 173-85.

Stoker, Laura, and Jake Bowers. 2002. Designing multilevel studies: Sampling voters and electoral contexts. *Electoral Studies* 21:235-67.

White, Halbert. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4): 817-30.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.