

Phoning It In: Overcoming Implementation Challenges in Field-Experiment Partnerships

Brian Robert Calfano, *University of Cincinnati*

ABSTRACT

The use of field experiments in political science has become extensive, but the promise of conducting a randomized intervention in a “real world” setting also raises perils for researchers. Partnering with organizations to deliver a randomized intervention may be a cost effective route to data collection, but a long-distance partnership presents certain challenges. In particular, the researcher needs to be especially vigilant about treatment application given the potential for noncompliance with the random assignment schedule. I provide an evaluation of a field intervention’s effectiveness where the long-distance partner organization inadvertently violated the random assignment of voter precincts in a canvassing effort prior to a citizen vote to repeal a non-discrimination housing ordinance protecting LGBT residents of a Midwestern city. I then provide recommendations for researchers to help mitigate treatment noncompliance when they cannot be present during treatment delivery.

The limitations of data generated through observational (e.g., survey) and experimental research in artificial settings such as laboratories often compel scholars “to allocate resources to field experimentation” (Gerber and Green 2008; Gerber, Green, and Kaplan 2014, 24; Green, Calfano, and Aronow 2014). Financial requirements associated with executing field designs, however, mean that scholars may have to seek cost-cutting approaches to use the methodology. An obvious option is to utilize the financial and personnel resources of a third-party organizational partner.

Yet, a partnering organization—although interested in a research idea—may be unwilling to relinquish substantial control of its personnel and imprimatur unless it maintains a dominant role in implementation. Moreover, with recent negative press from the Montana mailer and LaCour scandals, even scholars with national reputations may have difficulty establishing enough trust with a collaborating organization to work without substantial oversight. There also are potential partnership perils, particularly if researchers must “phone it in” because they cannot be physically present during the data-collection phase. This requires researchers to think through their actions when more limited implementation control threatens a study’s scientific validity.

Brian Robert Calfano is assistant professor of political science and journalism at the University of Cincinnati. He can be reached at Brian.Calfano@uc.edu.

The following sections (1) briefly describe a long-distance researcher–organization partnership in conducting pro-LGBT canvassing in advance of an April 2015 public referendum on a housing nondiscrimination ordinance in a Midwestern city; (2) discuss (unintended) noncompliance in treatment application; (3) estimate treatment effects following Nickerson’s (2005) protocol when noncompliance occurs; and (4) recommend what researchers can do to avoid these design issues when partnering long distance with an organization to conduct field interventions.

CANVASSING IN SUPPORT OF LGBT NONDISCRIMINATION

My organizational-partnership example concerns a canvassing effort to persuade voters to oppose the proposed repeal of a nondiscrimination housing ordinance¹ protecting members of the LGBT community. The repeal proposal was in the form of a public question on the municipal ballot used for an election in Springfield, Missouri, on April 7, 2015. I am the first researcher to assess canvassing effects when an LGBT issue was contested in an election setting.

The canvassers were recruited, trained, and remunerated by the Human Rights Campaign (HRC) and its local-affiliate partners. The HRC canvassers used a different approach than canvasser-based revelations of personal struggles as LGBT community members. The technique they used was pioneered by the Los Angeles LGBT Center and adapted for a Florida

study of attitudes toward transgender persons (Broockman and Kalla 2016). The HRC canvassers explained to voters that they had close friends and family members in the LGBT community who would be adversely affected if local legal protections such as the nondiscrimination housing ordinance were repealed.

I was referred to the HRC as a potential canvassing consultant based on research that I conducted with another third-party organization in the Springfield area a few years earlier. However,

before canvassing began because their boundaries were erroneously listed as within city limits. A total of 27 precincts remained on the master list (i.e., 14 treatment and 13 control).

CANVASSING AND REPEAL

Canvassing by HRC affiliates began on March 18, 2015, for households assigned to the treatment precincts. As efforts on both the “repeal” and “no repeal” sides ramped up in the 10 days leading to Election Day, the HRC affiliates added more

The HRC canvassers explained to voters that they had close friends and family members in the LGBT community who would be adversely affected if local legal protections such as the nondiscrimination housing ordinance were repealed.

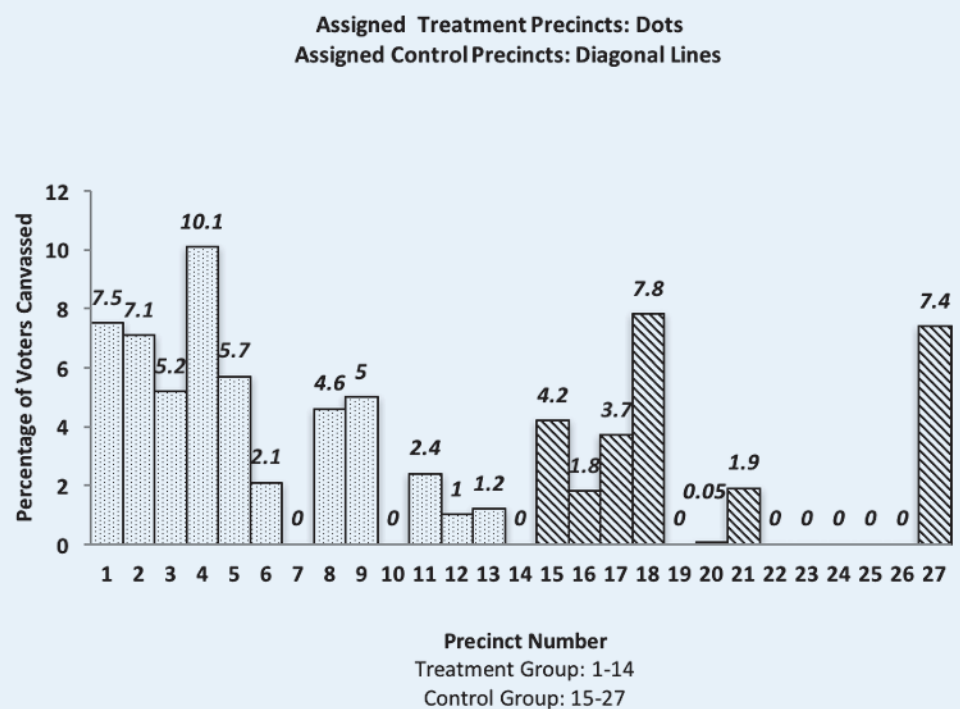
because I could not be physically present during the HRC canvassing, my input had to be provided long distance. After my initial telephone conversation with local HRC leaders, they agreed to allow the canvassed precincts in their campaign to be determined at random. By the time of my involvement with the canvassers, local HRC organizers had already trained their canvassing teams using proprietary organization materials, which described what the organization determined were the best techniques for sharing experiences of LGBT friends and family members.

My agreed-on role was to devise a randomization protocol for the canvassing teams to follow. Initially, the HRC’s canvassing resources allowed for a limited effort that could not target even half of the city’s precincts. Therefore, I block-randomized 31 of the city’s 72 precincts encompassing between 1,000 and 3,000 registered voters and that had shown 65% to 75% support for Missouri’s 2004 constitutional amendment banning gay marriage.

Each of the 31 precincts in this “between-subjects” design was assigned a priority number using a random-number generator. These numbers (and their corresponding precincts) were sorted from smallest to largest. The first 16 precincts on the list were assigned to the treatment, where they would receive pro-ordinance (i.e., “no repeal”) canvassing. I instructed canvassing teams to visit the treatment precincts in the order that they appeared on the list. The remaining 15 precincts were assigned to the control group and were not to receive any canvasser contact. Two precincts assigned to the treatment precincts and two assigned to the control group were eliminated

canvassing teams (as more donations poured into the “no repeal” effort). Perhaps in the enthusiasm of canvassing with these swelled ranks and resources, some canvassers visited households in seven precincts listed as control. (Canvassers did not visit any precinct that was not included on the master list of 27 treatment and control precincts.) Canvassers did not visit households in three of the 14 treatment precincts. Given my absence during the data-collection process, I was informed of noncompliance in the treatment application and control-group contamination after the April 7 vote. In the assigned-treatment group, canvassers contacted an average of 76 people per precinct; in the control group, they contacted an average of 26 people per precinct. Figure 1 shows the distribution of the percentage of voters canvassed per precinct.

Figure 1
Percentage of Voters Canvassed Per Precinct



The effort to repeal the citywide nondiscrimination ordinance was successful: 51.4% of voters (15,364) opted to repeal, thereby removing housing protections for LGBT residents, and 48.6% (14,510) voted to keep the ordinance. In the assigned-treatment precincts, “no repeal” votes averaged 48%; in the control-group precincts—including those contaminated in the canvassing visits—they averaged 43%. Figure 2 shows the distribution of the percentage of “no repeal” vote totals.

A purist perspective suggests that noncompliance in treatment application sufficiently threatens an experiment’s internal validity to require a do-over. However, casting aside the

effectiveness on precinct-level vote outcomes is worth an attempt.

TREATMENT EFFECTS IN LIGHT OF NONCOMPLIANCE

Similar to the use of Vietnam draft-lottery numbers to estimate outcomes based on a random assignment that is independent of individual characteristics (Angrist 1990), using the random priority number to predict the endogenous voter contact rate by the HRC canvassers—which, in turn, affected vote shares in the April 2015 election—enabled a test of a treatment’s statistical significance even in the face of noncompliance. Because the random

If I had simply abandoned the experiment because of noncompliance, the insights regarding the effects of canvassing about a controversial issue in the context of a local election would have been lost.

design and data is a costly decision for the researchers, partnering organization, and interested constituencies—especially when the experiment is conducted in an election context. Following Nickerson’s (2005) example, a suitable approach is to consider whether canvassers followed the randomization instructions well enough to enable the recovery of meaningful effect estimates. The fact that the majority of control precincts were not contaminated suggests that assessing the canvassers’

priority number was a scaled-back version of the original randomized treatment and control assignment, it should be related only to the vote shares.

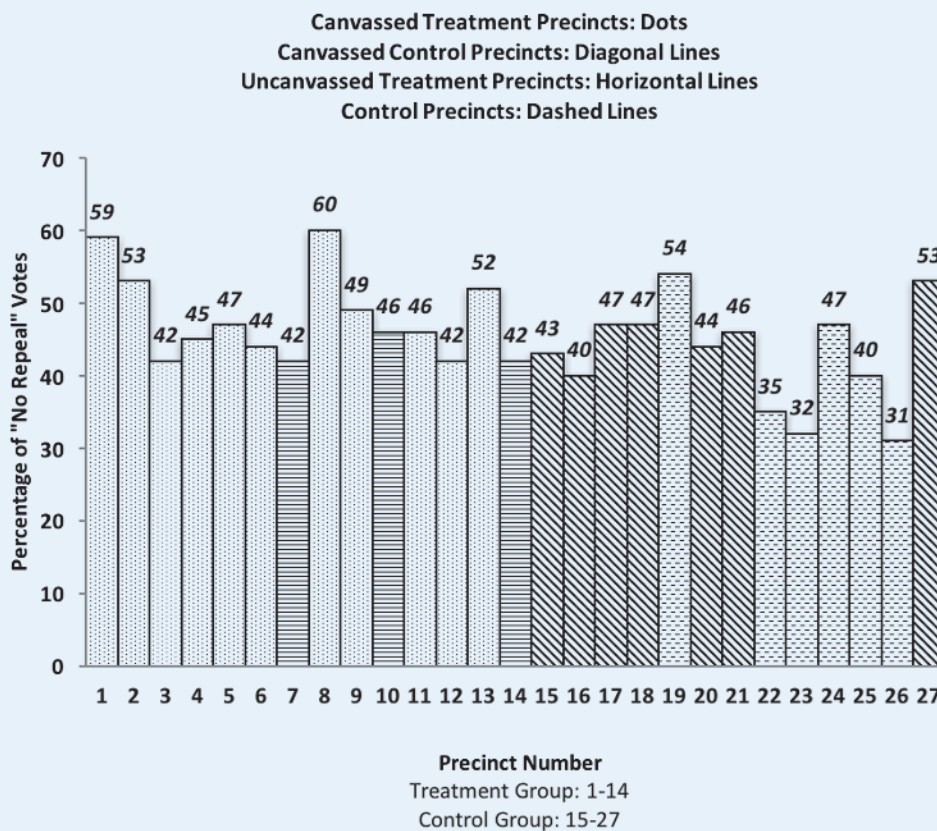
Nickerson (2005) provided an in-depth example of the random-priority-number approach, which he characterized as a “rolling experimental protocol.” In following this protocol, I measured the relationship among the random priority number—which was the instrumental variable (numbered 1 to 27, assigned to each precinct)—the voter contact rate by HRC canvassers per precinct, and the precinct level “no repeal” vote share (i.e., the two dependent variables).

Table 1 shows the relationship among the 1 to 27 randomly assigned priority numbers for each precinct, the voter contact rates in each precinct, and the shares of “no repeal” votes. Coefficients in table 1 indicate that as the probability of treatment increased (measured as the random priority number), so did both the precinct-level contact rate and the “no repeal” vote proportion.² Figure 3 plots the relationship between the random priority number and the precinct “no repeal” vote proportion.

The two blocking covariates—the number of registrants in each precinct and the share of votes cast in the 2004 statewide same-sex-marriage ballot measure—are included in models 2, 3, 5, and 6 in table 1. Both covariates were mean-centered so that when interacted with the random priority number, the priority number’s main effect corresponds to the

Figure 2

Percentage of “No Repeal” Votes Per Precinct



estimated effect. Lin (2013) advocated the use of this approach on the grounds of statistical efficiency. As model 6 shows, the precinct-level “no repeal” vote increased with the probability of treatment. Specifically—and with controls for interactions among random priority number, voter registration, and vote share included—the precinct-level “no repeal” vote increased by about 5.4 percentage points as the random priority number increased from 1 to 27.

Using the rolling experimental protocol, I recovered estimates showing an effect from the canvassing on the “no repeal” vote

1. Assume That Problems Will Arise in Long-Distance Designs *Prima Facie*

The responsibility for ensuring that a canvassing field experiment follows the proper protocols begins and ends with the researcher, not the partnering organization. The HRC’s goal was to reach as many households as possible to persuade voters about retaining the housing ordinance. My goal was to leverage the canvassing to generate insights on field interventions occurring during an election in which an LGBT issue

The researcher should always be sensitive to interpersonal dynamics between leaders and organization workers, but it is beneficial to develop an alternate point of collaborative contact within the partner organization.

proportion. If I had simply abandoned the experiment because of noncompliance, the insights regarding the effects of canvassing about a controversial issue in the context of a local election would have been lost. Of course, an even better outcome would have been to avoid noncompliance in the first place. Given, then, the long-distance nature of my partnership with HRC, I could have done a few things differently to ensure that the effect from any breakdown in protocols was mitigated.

PARTNERING RECOMMENDATIONS

The primary complication contributing to the noncompliance was the fact that I could not be physically present to manage the canvassing efforts during the treatment period. Although it is tempting to recommend that researchers always be present with their organizational partners in the field, this expectation may be unrealistic. Based on my experience, I offer the following recommendations for long-distance partnerships.

was on the ballot. Particularly because I was providing long-distance guidance to the HRC, I should have anticipated from the outset the possibility of a breakdown in communication regarding the importance of following the random-assignment schedule.

Better yet, I should have assumed that following the randomization schedule would not be executed consistently and tailored my interaction with the HRC accordingly. This posture is not meant to signal pessimism about a partner’s ability to follow the randomization schedule. Rather, it reflects the reality that starting a project with the mindset of troubleshooting and correcting deviations at every communication point with the long-distance partner provides the best opportunity to mitigate threats to treatment compliance. In being more proactive, I could have expanded the discussion about random assignment of precincts to the treatment and control groups to include as many of the actual canvassers as possible

Table 1

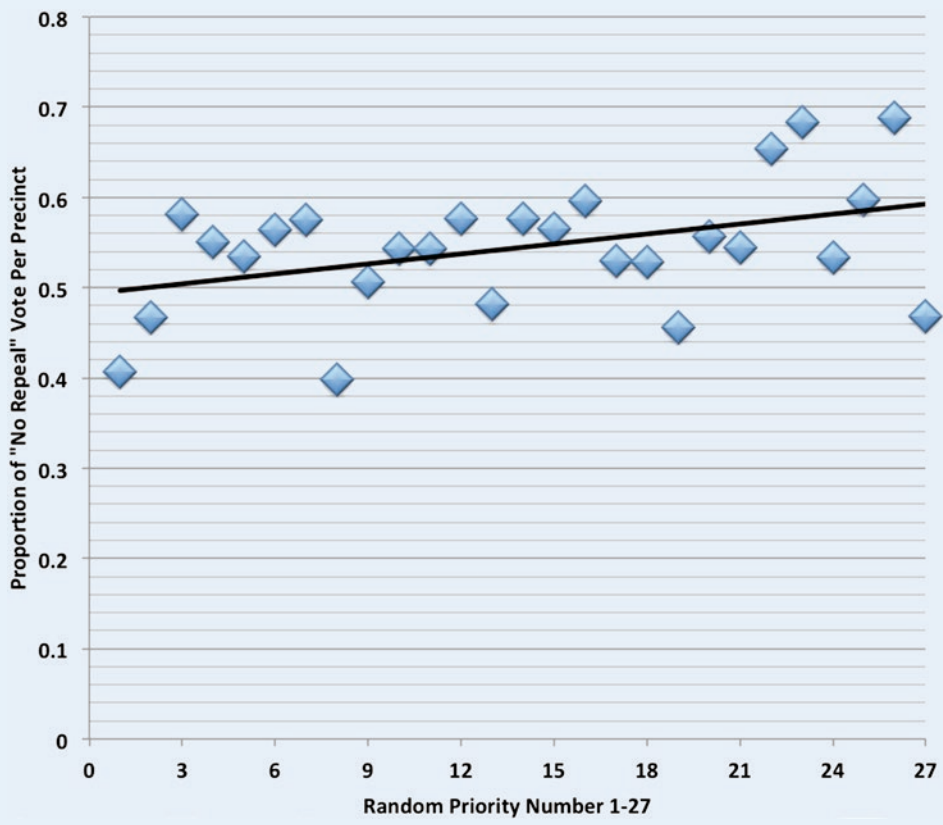
Effects of Random Priority Number on Contact Rate and Vote Share

	Contact Rate			Vote Share		
	Coeff. (Robust SE)					
	I	II	III	IV	V	VI
Random Priority Number	0.19 (0.08**)	0.18 (0.07**)	0.17 (0.06**)	0.37 (0.19*)	0.19 (0.12)	0.20 (0.11*)
Registration Control		0.003 (0.001**)	0.004 (0.002**)		0.002 (0.002)	0.005 (0.004)
Past Vote Share Control		-1.9 (17.05)	-47.64 (45.92)		116.29 (33.36**)	-9.71 (89.68)
Random Priority Number* Registration			9.59e-07 (1.33e-06)			1.39e-06 (2.56e-06)
Random Priority Number* Past Vote Share			2.70 (2.50)			7.83 (4.77)
Constant	5.6 (1.2)	5.4 (1.00)	5.51 (0.940)	50.69 (2.72)	48.22 (1.80)	48.97 (1.67)
N	27	27	27	27	27	27
R²	0.03	0.48	0.52	0.17	0.59	0.65

* = $p < 0.05$ (one-tailed) ** = $p < 0.01$ (one-tailed)

Figure 3

Random-Priority-Number Effect on Proportion of “No Repeal” Vote Share



3. Use Technology (and Persistence) to Engage Partners in Frequent Data Updates

It is tempting to arrange for organizational partners collecting data to provide this information at the end of design implementation. In fact, organizations may prefer this because it means less work than providing regular updates. The problem for researchers, however, is that there is an increased possibility for reporting error when providing a single data “dump” as well as fewer opportunities to communicate with organizational contacts about design protocols. I could have explored ways to make the data-update process more frequent and reliable without undue burden on the HRC. One possibility would have been to use online-survey research platforms programmed to “ping” organization workers to enter data updates that researchers can then download and track. Sharing a Dropbox folder or using Google sheets in which canvassing teams can frequently upload and revise information that is accessible to all is a way to gather data at sufficient intervals to ensure good monitoring. Alternatively, I could

(not only the team leadership). Creating an e-mail list of canvassers likely would have improved the chances of effective communication about the study and the intent of the random assignment. Furthermore, I could have tailored my conversations with the canvassers during the treatment period to include frequent reminders of the need to adhere to the randomized precinct list. In reality, I was concerned about sounding like a “broken record” regarding the random assignment in my discussions with HRC personnel. In hindsight, this was a risk worth taking.

2. Develop Easy-to-Use Reference Guides on Experiments for Organization Workers

Related to the previous point—and because it is beneficial to have organizational partners understand the logic behind experimental design—I also should have included materials on randomization as a supplement to canvasser training. For example, I could have created easy-to-use reference materials that explain the basics about why randomization is important for causal inference and the insights that a specific field experiment will generate for an organization and its constituents. The information could have been distributed as a pamphlet to field workers, placed on a website, or as part of smart phone app. Publicly available materials from organizations such as the Abdul Latif Jameel Poverty Action Lab would have been suitable for this purpose.

have offered to be in regular contact with representatives from each canvassing team via telephone or e-mail to collect their data individually. Whatever plan is chosen, be persistent in adhering to it.

4. Encourage Detailed Recordkeeping

As a corollary to the previous point, detailed recordkeeping—in addition to frequent data updates—is beneficial. Organizations have different degrees of interest in keeping records of their canvassing efforts, particularly in terms of the number and types of contacts made at residences during canvassing and Get Out The Vote efforts. It is useful to have conversations and make detailed plans about recordkeeping procedures and best practices before beginning any canvassing or data-collection efforts. Be prepared to provide templates, tutorials, and other support materials to organizational partners as part of this process. It also is helpful to determine whether the partner will agree to a random audit of a certain percentage of the canvassing and data-collection work at regular intervals during a campaign. The rationale in explaining this request should be stated as ensuring quality control for the benefit of achieving the organization’s goal in the canvassing work.

5. Develop Alternate Collaboration Points with Organizational Partners

The researcher should always be sensitive to interpersonal dynamics between leaders and organization workers, but it is beneficial to develop an alternate point of collaborative contact

within the partner organization. The goal concerns more than having another contact person to discuss the status of the data collection. With broader discussions about the insights generated from field interventions with multiple staff members in the partner organization—particularly those who can reinforce the researcher’s message to local canvassing and data-collection teams in the field—the chances for greater organizational buy-in to maintain the randomization schedule increase.

At the same time, the reality is that it may be difficult to have frequent contact with a single person in the organization during the course of a research campaign. Invariably, some leaders will be better organizers and others will be better communicators. In a case in which researchers are working with a person who is not good at both skills, they should explore the option of widening the circle of collaboration within the organization. Couch the request as a way to take logistical pressure off of the immediate organizational contact by communicating with others in the organization who are involved in the canvassing, data-collection, and related activities.

6. Plan Designs before You Are Asked to Partner

Gaming out different types of field interventions in the absence of an actual opportunity to do an experiment may seem like a waste of time. However, if I had had an existing plan that provided greater clarity of objectives in the initial hours and days after being approached to work on the Springfield canvassing project, I might have been better able to troubleshoot where breakdowns in communication and protocols were likely to occur. I would have been better prepared to encourage the HRC to do whatever was necessary to ensure that the canvassers understood the importance of adhering to the randomization schedule—even as the extra resources accumulated. Developing research designs without actually conducting data collection and analysis is a throwback to graduate school—and this is not a bad thing.

7. Be Prepared If Partners Change Their Mind about a Randomized Intervention

My working assumption—and what the HRC indicated after the canvassing concluded—was that the noncompliance was a matter of oversight. However, what if the HRC (or another canvassing-organization partner) decided in the middle of the treatment application that it no longer supported the idea of randomly treating only some precincts and stopped following the randomization schedule as a result? This would be a more complex situation than what I encountered, but a partner’s changing perceptions and needs during the course of data collection requires the researcher to have honest conversations about these concerns regarding the use of randomized interventions. It also requires

the researcher to accede to a partner’s wishes if there is a change of mind about conducting the field experiment.

8. Consider Saying No

It never occurred to me to say no to the HRC’s request. However, researchers should be open to the idea that although an organization makes a collaboration request, the challenges of a long-distance partnership may make working together an unsatisfactory experience. Through these recommendations, it is hoped that long-distance collaborations on field experiments might flourish. Nevertheless, scholars should not hesitate to take a sober inventory of their ability to work closely with partner organizations to ensure design validity.

ACKNOWLEDGMENT

I thank the anonymous reviewers for their helpful comments in preparing this article. ■

NOTES

1. Ordinance Language: “Shall the City of Springfield, Missouri, repeal General Ordinance No. 6141, adopted by City Council on October 13, 2014, that amended Chapter 2, Administration, Article IV, Boards, Commissions, and Committees, Division 4, Mayor’s Commission on Human Rights and Community Relations, Section 2-223; and Chapter 62, Human Rights, to add sexual orientation and gender identity to the list of categories of persons protected from discrimination, to consolidate the Commission’s investigative process into a single division of the Code, and to clarify the Commission’s powers in light of recent court decisions; and readopt the City’s prior Code of Ordinances regulating Chapter 2, Administration, Article IV, Boards, Commissions, and Committees, Division 4, Mayor’s Commission on Human Rights and Community Relations; and Chapter 62, Human Rights?”
2. The original direction of the coefficients for the expected effect in the models was negative. For ease of interpretation, I reversed the signs so that positive coefficients represent an increase in support for the nondiscrimination ordinance—that is, the “no repeal” vote.

REFERENCES

- Angrist, Joshua D. 1990. “Lifetime Earning and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.” *American Economic Review* 80: 313–36.
- Broockman, David, and Joshua Kalla. 2016. “Durability Reducing Transphobia: A Field Experiment on Door-to-Door Canvassing.” *Science* 352: 220–24.
- Gerber, Alan S., and Donald P. Green. 2008. *Get Out the Vote: How to Increase Voter Turnout, second edition*. Washington, DC: Brookings Institution.
- Gerber, Alan S., Donald P. Green, and Edward H. Kaplan. 2014. “The Illusion of Learning from Observational Research.” In *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, ed. Dawn Langan Teele, 9–32. New Haven, CT: Yale University Press.
- Green, Donald P., Brian R. Calfano, and Peter M. Aronow. 2014. “Field Experiment Designs for the Study of Media Effects.” *Political Communication* 31: 168–80.
- Lin, Winston. 2013. “Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique.” *The Annals of Applied Statistics* 7: 295–318.
- Nickerson, David W. 2005. “Scalable Protocols Offer Efficient Design for Field Experiments.” *Political Analysis* 13: 233–52.