Do Nonpartisan Programmatic Policies Have Partisan Electoral Effects? Evidence from Two Large Scale Randomized Experiments*

Kosuke Imai†  Gary King‡  Carlos Velasco Rivera§

February 14, 2016

Abstract

A vast literature demonstrates that voters around the world who benefit from their governments’ discretionary spending cast ballots for the incumbent party in larger proportions than those not receiving funds. But contrary to most theories of political accountability, the evidence seems to indicate that voters also reward incumbent parties for implementing “programmatic” spending legislation, passed with support from all major parties, and over which incumbents have no discretion. Why voters would attribute responsibility when none exists is unclear, as is why minority party legislators would approve of legislation that will cost them votes. We address this puzzle with one of the largest randomized social experiments ever, resulting in clear rejection of the claim that programmatic policies greatly increase voter support for incumbents. We also reanalyze the study cited as claiming the strongest support for the electoral effects of programmatic policies, which is also a very large scale randomized experiment. We show that its key results vanish after correcting either a simple coding error affecting only two observations or highly unconventional data analysis procedures (or both). We also discuss how these consistent empirical results from the only two probative experiments on this question may be reconciled with several observational and theoretical studies touching on similar questions in other contexts.

*We are grateful to Ana De La O for data and replication information, and for help with followup questions; to Isadora Antoniamo (formerly of the Mexican Electoral Institute), Miguel Rojano (Director of Electoral Cartography) and Luis Ruvalcaba (Deputy Director of Geographical Electoral Systems Development) for discussing with us all the intricacies of electoral cartography in Mexico; and to Rikhil Bhavnani, Graeme Blair, Chris Blattman, Ken Greene, Guy Grossman, Macartan Humphreys, John Londregan, Will Lowe, Horacio Larreguy, and Grigore Pop-Eleches for helpful comments.

†Professor, Department of Politics and Center for Statistics and Machine Learning, Princeton University, Princeton NJ 08540; http://imai.princeton.edu, kimai@princeton.edu, (609) 258-6601.

‡Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge MA 02138; GaryKing.org, King@Harvard.edu, (617) 500-7570.

§Ph.D. Candidate, Department of Politics, Princeton University, Princeton NJ 08540; http://www.cvelasco.org, cvelasco@princeton.edu.
1 Introduction

Political scientists in American and comparative politics have amassed considerable support for the theory that office-holders target discretionary government spending to gain votes, and voters reward them for doing so. As the exhaustive literature review by Golden and Min (2013, p.12) summarize: “Studies overwhelmingly find that incumbent politicians are rewarded by voters for distributive allocations, and in particular for those that are clientelistic and from which recipients can be excluded.” However, aspects of this well developed literature are called into question by other results we discuss below claiming to show that the non-discretionary spending required by programmatic policies (see Kitschelt and Wilkinson, 2007, for a definition) also motivate voters to cast their ballots for the incumbent party. This is all the more puzzling because programmatic policies are usually passed with support from every major political party, including those which, according to the literature, will be hurt electorally by their very action.

Consider, for example, the case of conditional cash transfer programs in Brazil and Mexico, the basis for many of the studies of the electoral effects of programmatic policies. Here, voters knew that such programs would be in place regardless of which party held office. Prior to the 2006 Brazilian election, even though one set of incumbents were in office, all candidates spanning the left-right spectrum supported the expansion of the program (Zucco, 2013, p.819). Similarly, voters in Mexico knew the benefits of its program would be automatic and thus no longer subject to the whims of the party in power (Diaz-Cayeros, Estevez and Magaloni, Forthcoming, ch.6). Moreover, although this program was of course implemented by one party, it was supported on passage and afterwards by all political parties; indeed it remained in place despite changes in party control of both congress and the executive (Levy, 2006, p.108).

This, then, is the puzzle we address: Why would minority party legislators support bills that will reduce their own support in the next election? And why would voters reward a party for actions over which the party has no control? Although some efforts have been made to address these questions via qualitative argument (Cornelius, 2004; Diaz-Cayeros, Estevez and Magaloni, 2009, Forthcoming), and via the possibility of publication bias
We address this puzzle by analyzing two large scale randomized experiments. The first, which we designed and implemented originally for the purpose of evaluating *Seguro Popular de Salud* (SPS) in Mexico, is one of the largest social experiments ever, and the largest randomized health policy experiment to date (King et al., 2007). The second is the only randomized social experiment previously used to study the electoral effects of programmatic policies, and the one regularly cited as giving the strongest support (De La O, 2013, 2015). Like the first experiment, it was also very large, originally designed as a policy evaluation, and in Mexico. The policy it evaluated was the *Progresa* (later called *Oportunidades*) anti-poverty program.

The policies in both experiments (a) were passed with broad support in the legislature from all major parties, (b) came with objective rules that gave incumbents no discretion over implementation, (c) did not lead any single party to attempt to claim exclusive credit for its implementation, and (d) had voters expecting to receive the same services regardless of the party in office. After introducing the design and policy background of the two experiments (Section 2), we present our results (Section 3), which are consistent and unambiguous: *we find no evidence that nonpartisan programmatic policies effect electoral support for the incumbent party.* We then discuss how our results may fit into the broader empirical literature where all four policy conditions do not always apply, and into the formal literature attempting to explain how it all works (Section 4). We also provide an Appendix that provides technical information about merging census and electoral data, and a separate online Supplementary Appendix offers extensive supporting statistical analyses, robustness checks, alternative measurement strategies, and other information.

## 2 A Tale of Two Experiments: Background and Design

We now briefly describe the two huge programmatic policy interventions and the experiments designed to evaluate each that we marshal for this study. The facts that these programmatic policies were large, and the experiments we now describe had large ef-
effects on their intended social policy outcomes, makes them exceptionally hard tests of our hypothesis that they do not increase incumbent support.

**Seguro Popular** Although *Seguro Popular de Salud* (SPS) translates literally to “universal health insurance,” the Spanish word for “insurance” does not appear in the authorizing legislation, as it is a social welfare (income redistribution) program, not a self-sustaining insurance program. The program was designed to build or improve medical facilities, and provide medical services, preventive care, pharmaceuticals, and financial health protection to the 50 million Mexicans with no regular access to health care (constituting about half the population of the country). SPS was aimed at those with low incomes; its main purpose is to reduce the devastating effects of catastrophic health expenditures, when, due to illness or injury, greater than 30% of a family’s annual disposable income is spent on health care in one year. Before SPS, about 10% of the poor had catastrophic health expenditures each year. SPS was designed to eventually spend an additional one percentage point in GDP; in 2005, expenditures totaled a substantial $795.5 million. As it turned out, SPS was the most visible accomplishment of the administration of President Vicente Fox Quesada. By all accounts, it was designed, passed, and implemented in a nonpartisan, programmatic fashion.¹

Because of Mexico’s term limits, President Fox, along with Health Minister Julio Frenk Mora, decided they needed a way to convince whoever would succeed them to keep SPS in place. How one democratically elected government can “tie the hands” of the next democratically elected government is a classic question of normative political theory (Klarman, 1997; Posner and Vermeule, 2002; Sterk, 2003; Thompson, 2005), formal theory (Alesina and Tabellini, 1990), and empirical political science (Franzese, 2002). In this case, Fox and Frenk’s approach was to commission an independent scien-

---

¹Julio Frenk, one of SPS’s main architects said “From the beginning I suggested that this [SPS] had to be a proposal from all parties, in other words, that it was not only a PAN, PRI, or PRD project, but a project for the country, for which everyone could claim credit. In a conversation with President Fox, after introducing the bill to Senate, I told him: ‘If we want the reform to move forward, it is extremely important that the project is handled as a shared project and that we give credit to everyone.’ The president agreed, and promised that credit would be given to everyone, which is what happened” (Ortiz, 2006, p.81, our translation). Whether we judge by public statements from minority parties, or from the votes of legislation, there exists near consensus on this point.
tific evaluation by a team we led at Harvard’s Institute for Quantitative Social Science (see http://j.mp/ExpMex). The idea was that if the evaluation favored SPS, it would at least have been more difficult for the next government to eliminate. And to make the evaluation worthwhile, we committed publicly and in print to say so if the evaluation was not favorable in any way.

The Mexican government signed legal contracts that gave us free rein to design and implement whatever evaluation, and spend whatever funds, we judged appropriate. The government gave us unfettered access to government officials, the ability to influence how SPS was implemented so we could more easily evaluate it, and convening power to speak with the numerous local officials across the country in charge of implementation. We retained the legal right to publish without prior review.

We developed a new experimental research design robust to the interventions by politicians who regularly — indeed usually — derail large public policy experiments, as they choose to be more attentive to the short term desires of their constituents than any longer term benefit of scientific evaluation (King et al., 2007). Details of the experiment include discussions of the background and design, published prior to any data analysis (King et al., 2007), novel statistical methods we developed (Imai, King and Nall, 2009a), empirical results (King et al., 2009a), and publicly available replication data (Imai, King and Nall, 2009b; King et al., 2009b).

We began by defining 12,284 health clusters, which are new, continuous geographic areas we defined that tile all of Mexico’s 31 states in which fall one health clinic (or potential clinic that we could decide would be build) along with its catchment area (defined as less than a day’s travel time to the clinic, using locally available methods of transportation rather than “as the crow flies”). We persuaded 13 of Mexico’s 31 states to participate, including 7,078 clusters. We matched these clusters in pairs based on background characteristics and then selected 74 pairs (based on the closeness of the match, likelihood of compliance with the experiment, and necessary political criteria). The experiment thus included 534,457 research subjects, in 118,569 households, within 1,380 localities, nested within 148 health clusters. Within each of the 74 matched pairs, we randomly selected
one cluster to receive SPS and the other, as the control, to receive no change. Health clusters which we randomly assigned to the treatment group received new (or upgraded) hospitals and other medical facilities, doctors, access to medicines and other medical interventions, advertising campaigns to encourage affiliation with the program, and funds to pay for it all. The treatment was applied August-September, 2005, coincident with a baseline survey of 32,515 respondents randomly selected from 50 of the cluster pairs; the outcome was measured the same way 10 months later, July-August 2006.

To merge federal election results reported by one governmental office, with census data reported by another, we define a new unit of analysis for this study within the confines of the experimental randomization. This is the **precinct cluster**, which we define as the largest possible geographic subset of a single health cluster in the SPS experiment for which we can accurately merge all relevant electoral and census information (for details, see the Appendix). When insufficient information is available to identify a precinct cluster, we retained all the benefits of this matched pair cluster randomized design by removing it and the health cluster (and corresponding precinct cluster) to which it was pair matched (as suggested by Imai, King and Nall, 2009a). This left us in the end with 57 matched-pairs (47 rural and 10 urban) out of the original 74. Table 1 and Figure 2 in the separate online Supplemental Appendix present full details and descriptive statistics of the sample we use and to which our inferences apply. (As another robustness check, we repeated our analyses with all available precinct clusters even when no match was available, trading off more model dependence for less inefficiency, and found no change in the results we present below. See Tables 2 and 3 in the Supplementary Appendix.)

It is especially useful for the present paper that our evaluation indicated that SPS had a massive financial effect (a fact which is all the more impressive because government programs designed to help the poor in most countries typically have no measurable impact on the poor; Gwatkin, Wagstaff and Yazbeck 2005). SPS eliminated about a third of the catastrophic health expenditure problem among the poor in Mexico, and about 60% among experimental compliers (those who would affiliate to SPS when in the treatment group and not when in the control group). Some aspects of the program did not have

---

2This decision is not affected by which unit received the treatment within a given pair.
the intended effects, as we reported (King et al., 2009a), but for our purposes this large financial impact makes this an especially strong test of our claim for nonexistent electoral effects of programmatic policies.

**Progresa**  Our second experiment evaluates one of the largest poverty-alleviation programs in Mexico. The program consists of nutritional, educational, and health components. The policy’s key feature consists of cash transfers that eligible households receive on the condition that they attend regular health check-ups, and children enroll and attend school. Program benefits vary according to household composition. The average level of benefits is about 35 USD per-month, which represents about 25% of income in poor rural households (Levy, 2006, p. 23).

The origins of the program date back to the mid 1990s, when the country experienced one of the worst economic crises in its history. The government had previously relied on a myriad of food subsidies to alleviate poverty. However, government officials in the administration of then president Zedillo considered that subsidies benefitted urban centers at the expense of the poorest rural areas in the country, and were regressive and too costly to administer (Levy, 2006, ch. 2). The motivation underlying the new administration’s alternative approach was a recognition that in order to break the cycle of poverty, one had to recognize the relationship between the educational, health and nutritional components of human capital.

The architects of the program were committed to the eradication of poverty in the country, and as a result they implemented a design that would increase its long-term viability. In contrast to previous government policies, Progresa defined a target population and operated under clear, programmatic rules. In the eyes of the policy-makers, this would ensure the political neutrality of the policy.³ Political scientists have also arrived at the

---

³Santiago Levy, one of Progresa’s main architects, emphasized the nonpartisan goal of the program: “Congress role in Progresa-Oportunidades has also contributed to its continuity in yet another way: it has established strong legal provisions against the ‘political’ use of the program. More particularly, it has sought to separate the program from the public image of the president and to provide information directly to beneficiaries about the nature of the benefits that they receive, their rights, and their obligations…. These factors, along with the programs positive results, have contributed to the programs transit through three shifts in the composition of the House of Representatives since 1998” (Levy, 2006, p.107-108).
same conclusion about the nonpartisan nature of Progresa.\footnote{Diaz-Cayeros, Estevez and Magaloni (Forthcoming) conducted interviews across communities in Oaxaca – a state with a long history of clientelism – about the experience of voters regarding the provision of Progresa. In their interviews, one voter noted that “One can be PANista, PRIista or PRDista and still receive benefits from Oportunidades … before you had to be with the PRI to get anything from the government.” Another respondent noted that “the governor controls everything in Oaxaca. However, here you can be PANista, be with the governor [back then from the PRI], and still get benefits from Oportunidades.” A final interviewee asserted that “although sometimes people who do not really need it get Oportunidades, it is less corrupt because benefits arrive regardless of which party you like.” On the basis of evidence like this, Diaz-Cayeros, Estevez and Magaloni (Forthcoming, p. 195) conclude: “In our interviews in villages in Oaxaca, it became clear that the poor perceive big differences between Oportunidades and other social programs, and that they are generally most satisfied with the former because they perceive it as an entitlement rather than a political favor that comes and goes according to the waves of elections.”}

Analogous to the experience with SPS, the government hired the International Food Policy Research Institute as a third-party entity to evaluate Progresa and bolster the program’s credibility (Levy, 2006, 43).

The main evaluation of Progresa exploited the phasing of the policy across the country’s localities. The sample consists of 506 rural localities in the country distributed across seven states, with 320 villages drawn from a population of localities eligible to receive the program by November 1997. Eligible localities were then assigned to the treatment group. The remaining 186 villages were randomly drawn from a population that would receive the policy in one of the later phases, and thus constituted the control group. Then, the government carried out several surveys in each of these villages, first to determine household eligibility and then to measure outcomes over a period of two years. Behrman and Todd (1999) find that treatment and control villages are fairly similar across a large battery of socio-economic indicators.

The results of the evaluation based on this sample of villages shows that Progresa increases school enrollment, improves the health and nutrition of children and adults, and increases household consumption (largely on food) (Skoufias, 2005, Ch. 5).

As with the SPS experiment, analyzing federal election outcomes in Mexico requires a procedure for merging or matching the boundaries of electoral precincts with often overlapping census geography, as the two are generated by different administrative offices that typically do not coordinate. The same issue exists in almost all analyses of electoral data around the world, but the method of dealing with it is crucial. As we explain below, errors in this merging process for Progresa explains some of its unusual results.
These two extraordinary, large-scale experiments serve as an unusually powerful test of whether or not programmatic policies have electoral effects. Given the massive effects each of these policies had on people’s lives, if programmatic policies can electorally benefit incumbents, we should be able to find evidence for it in these experiments.

3 Results

The two sections that follow summarize our results. Numerous auxiliary analyses and robustness checks appear in the Supplementary Appendix that accompanies this paper.

3.1 Experiment 1: Seguro Popular de Salud

We now present two sets of outcome variables for the SPS experiment based on analyses of actual electoral results (using aggregation procedures described in the Appendix) and on retrospective survey evaluations (which require no aggregation), respectively. The statistical methods used in both cases are fully nonparametric, enabling us reap the benefits from our randomized experiment without making modeling assumptions (for a full description, see Imai, King and Nall, 2009a).

First, we analyze the causal effect of SPS on electoral outcomes in Figure 1. A point estimate (black dot) and 95% confidence interval (vertical line) appear for each causal effect, with a horizontal line at zero, indicating no effect. If the literature’s hypothesis is correct, that nonpartisan programmatic policies increase support for incumbents, the points and confidence intervals would appear above the horizontal line. Instead, all the confidence intervals cross the zero, no effect, line, and thus none confirm the hypothesis.

The overall estimated causal effect of SPS on the percent voting for the incumbent party (the PAN) is the first result, at the far left of Figure 1 (solid vertical line); the effect on federal voter turnout appears next to it (dashed vertical line). In both cases, the causal effect is not distinguishable from zero. Indeed, the confidence intervals are quite narrow for the estimated overall effects, indicating that the experiment is well powered for this hypothesis and so we should have high confidence that the effect is negligible if not exactly zero.
Figure 1: Causal Effect Estimates of SPS on Voter Turnout and Incumbent Party Vote. This figure gives point estimates and 95% confidence intervals for the “Intent-to-Treat” causal effect of SPS on voter turnout (solid line) and the incumbent (PAN) vote share (dashed line), overall (far left of left panel), and in three partitions of the data, including urban/rural (left panel), income quartile (middle), and asset quartile (right).

To further search for possible support for the hypothesis, we partition our sample in three different ways to examine subgroup effects. We give the causal effect within rural and within urban precinct clusters at the right side of the left panel. All results are statistically indistinguishable from zero. The right two panels estimate causal effects by quartile of the proportion of individuals in the first two income deciles (center) and household assets (right). These may be especially relevant since households that report less income will pay less (or nothing if in the first two income declines) for SPS services. These results can also be viewed as predicting compliance with receiving the experimental treatment, since less poor communities were expected to (and actually did) sign up for SPS services less often. This analysis follows Calvo and Murillo (2004), whose account suggests that the largest electoral effect of policies could be found among individuals who benefit most from a given policy. Nevertheless, the causal effect of SPS on voter turnout and on incumbent party vote within every one of these segments of the public is not distinguishable from zero.

Second, we avoid all issues involved in aggregating precincts to precinct clusters (discussed in the Appendix) by estimating the causal effect of SPS on individual survey evaluations. We do this for how well respondents thought the country was doing on economic,
political, and social issues, compared to five years before (we provide descriptive statistics in Figure 3 of our Supplemental Appendix). This analysis also allows us to test the channels through which retrospective voting may work, one of the mechanisms hypothesized for why incumbents may benefit from programmatic politics (Pop-Eleches and Pop-Eleches, 2012).

These results appear in Figure 2 (in a format parallel to Figure 1). For the overall effect (at the far left) and for each of the three domain areas (economic as a solid line, political as dashed, and social as dotted), the effect is estimated at near zero with a confidence interval that overlaps zero. We find the same essentially zero effect of SPS, and failure of the programmatic policy hypothesis, for individual effects within different subgroups, including rural, urban (left panel), income quartile (center panel), and asset quartile (right panel).

![Figure 2: Causal Effect Estimates of SPS on Retrospective Survey Evaluations.](image)

The figure reports point estimates and 95% confidence intervals of the causal effect of SPS on economic (solid vertical lines), political (dashed), and social (dotted) retrospective evaluations of whether the country was doing better today than it was five years ago \((n = 32,515)\) individuals in 50 matched health cluster pairs. Results are reported for all respondents and by urban/rural breakdown (left panel), income quartile (center), and asset quartile (right).

In our Supplementary Appendix, we also present a wide array of other analyses of the same data. For example, we reran Figures 1 and 2 with different coding rules and statistical techniques. These and many other analyses\(^5\) reveal no noticeable effect, which is also

---

\(^5\)To be specific, in the Supplementary Appendix, Figure 4 analyses PRI votes as a share of registered
a confirming check on the treatment randomization. Overall, the results are unambiguous: the highly successful SPS programmatic policy had little effect on turnout or the vote.

3.2 Experiment 2: Progresa

We are fortunate to be in the possibly unprecedented position for political science of having a second large scale randomized experiment to study the same question. Thus, we now use the Progresa experiment by replicating the analyses in De La O (2013), which were repeated and extended in a book De La O (2015).

We show here that the reported positive results in these works supporting the partisan effects of this nonpartisan programmatic policy were due to an unfortunate interaction between simple coding errors and highly unconventional data analysis procedures. When we correct either problem (or both), the results from this experiment look almost identical to those from our analysis of SPS in Section 3.1, clearly indicating little or no effect of programmatic policies on federal voter turnout or incumbent support. Although, in the end, we reach the unavoidable conclusion that the claims of this article and book are not supported by its empirical data, De La O deserves credit for highlighting this important issue, thinking of the idea of repurposing a randomized experiment to study it (as in Baldwin and Bhavnani, 2015), gathering the necessary data, and making available a replication data set so that these further discoveries became possible.

Coding Errors

We begin with the coding errors, the results of which are apparent in the first column, for voter turnout, and the third, for the incumbent (PRI) vote, in Figure 3. Since it is impossible for more than 100% of people to vote or to vote for any one party, every observation that appears above the dashed line at 100 is an error. Moreover, these include some large errors, cases that extend into the impossible region beyond 50% of the range and eligible voters, as alternative measures of incumbent support. Figure 5 measures turnout with total votes cast as a share of eligible voters. Tables 2 and 3 report estimates for the impact of SPS on incumbent support and turnout when we break up precinct clusters pairs and analyze the data in a regression framework. Tables 4 and 5 do the same but for the sample of rural precinct clusters, for which we know the population measurement error is minimal. Finally, Figure 6 reports estimates for the impact of SPS on individual retrospective economic, political, and social evaluations of the country when we control for the responses of the same people in a survey we conducted at baseline.
of the original data, and for turnout more than three times the range. To choose the extreme point, turnout obviously cannot be 375%, as it is in the data analyzed in De La O (2013, 2015). In addition to these known errors, the large number of observations with nearly zero turnout (left column, bottom) in a presidential election with about 65% overall turnout (IFE, 2013) would also seem to be of dubious validity.

Figure 3: **Distribution of Turnout and Incumbent Party Vote in the 2000 election.** This figure compares the variables originally constructed in De La O (2013) (in the first and third plots), with the official vote among registered voters (in the second and fourth), for both turnout (in the first and second) and vote for the incumbent party, the PRI (in the third and fourth).

Before we turn to the causes and consequences of these data errors, we analyze turnout and incumbent (PRI) vote among those officially registered (in columns 2 and 4 in Figure 3). These alternative variables are important for two reasons. First, they suffer from none of the problems in De La O (2013, 2015), which we describe below, because all the data needed to compute them come from the same source, with the same geographic boundaries. As a result, we can see in the figure that all the observations for these variables naturally fall within the possible region between 0 and 100%.

And second, the causal effect of Progresa is of course zero among all those not registered; yet, these are counted in the turnout and incumbent vote denominators in De La O.
(2013, 2015), even though they can never contribute to the numerator. In other words, infants, noncitizens, those not registered, etc., cannot vote. Except in the unlikely situation where these ineligible people randomly distribute themselves geographically, causal effects based on these variables can be biased. And even if they were randomly located, this coding strategy adds large and unnecessary inefficiencies in estimation.

Our alternative turnout and incumbent vote share variables, based on officially registered voters, have no such measurement error bias problem since all those in the denominator would appear in the numerator if they choose to vote. If Progresa works in part by increasing levels of registration — as hypothesized in De La O (2013, p. 7–8) — then the interpretation of the turnout and vote variables would change, but no post-treatment or other bias or inefficiency would be induced. Since no evidence has previously been offered for this hypothesis, we directly test it and present our results in Tables 12–15 of our Supplementary Appendix. The results, with small confidence intervals around zero, clearly demonstrate that the program had little or no impact on registration rates. For this reason, then, the alternative coding we offer for turnout and vote is a considerably cleaner test of the programmatic hypothesis, even if the original data had no coding errors.

Thus far, we have revealed the existence of data errors, but that alone is insufficient to change any conclusion; the real question is whether these errors make a difference in the results. We find that this is indeed the case, especially for two huge outliers. The particular impact of these outliers is their correlation with the control variable “lagged population,” which we graph in Figure 4. In the left panel, we graph the raw data, which reveals the two outliers (precincts 266 and 1502), each 90 to 100 times the size of the median and both of which appear in the treated group, with nothing comparable in the control group. De La O (2013, p. 6) was right to check balance between the means of the treated and control groups, but checking solely based on the means was a mistake that led to missing the massive imbalance in the tails of the two distributions evident in the figure. The consequence of imbalance is model dependence (Ho et al., 2007). Since the most important advantage of a correctly analyzed randomized experiment is the absence of model dependence, introducing these coding errors eliminated this important benefit of
randomization. Below, we illustrate and correct this problem.

Figure 4: **Population Outliers and Statistical Leverage.** This figure reveals the extreme degree to which two of the miscoded observations are outliers (left panel) with extremely high statistical leverage (right panel). The left graph is raw data; the right is computed from the least squares analysis in De La O (2013).

We also go a step further and formally evaluate the bias these (incorrect) outlier observations have on the least squares analysis in De La O (2013). We do this by computing the “statistical leverage” of these observations, with and without the lagged population variable. Observations with larger values of leverage have more influence on the magnitude of coefficient estimates a linear regression.\(^6\) As the right panel in Figure 4 makes clear, the statistical results in De La O (2013) are driven largely by these two observations.

Indeed, when controlling for lagged population, the two precincts have leverage that is 15 to 20 times larger than the median (see the vertical axis of the right panel). In contrast, when population is not included as a covariate, the leverage of these precincts is a modest 1.5 to 2.75 times larger than the median (see the horizontal axis). Since leverage is computed solely from the explanatory variables, this result applies identically to both outcome variables, turnout and incumbent vote. Particularly unfortunate is that both of these

\(^6\)The statistical leverage of observation \(i\) is defined as \(\frac{x_i^\top (X^\top X)^{-1} x_i}{n}\), where \(X\) is an \(n \times k\) matrix of pre-treatment covariates with \(k \times 1\) row \(x_i\).
extremely high leverage observations happen to also have extremely (and unrealistically) low reported turnout rates (1.78 and 2.59 percentage points, respectively) and PRI vote shares (0.74 and 1.08 percentage points). Judging from how the data were constructed, the considerably larger vote and turnout results based on official registration, the much larger vote and turnout national figures, or by comparisons with similar precincts, it is likely that these extremely low turnout and vote percentages are incorrect.

We have thus demonstrated all the conditions for bias in the main results: Substantial imbalance in the treated and control distributions, huge outliers highly correlated with a covariate included in the regression, and extreme outcome variable values for the same outlier units. We tracked these source of these problems to data errors resulting from the often difficult process of merging electoral and population data generated by separate government agencies that do not coordinate their work and have chosen different, overlapping geographies. For instance, while electoral authorities assign a given level of population to a precinct, census officials may aggregate part of the same population to a neighboring village centroid that happens to be outside that precinct. Neither office is necessarily correct or incorrect; but they have goals and methods that differ from each other. The problems due to lack of inter-agency coordination is a common problem in the analysis of elections in many countries, but the issue caused particular difficulties here when De La O (2013, p. 5) “overlayed the 506 experimental villages to the smallest unit of outcome measure for which census, precinct, and electoral data roughly coincide: the sección electoral (precinct).” The key problem is that De La O (2013) matched precincts by hand based on the textual names of villages in different data files from different organizations which assigned different meanings to the same names. We verify in the appendix to this paper that the matches in these cases were incorrect, with villages matched to areas with similar names from places far outside of designated treatment and control precincts. The problem was not due to carelessness in matching names; instead, the entire process of name matching in these data is unjustified because names of different areas that happen to match often do not correspond to the same geographical areas.

An important clue comes when De La O (2013, fn.17) tries to account for some of
the data with turnout and vote percentages greater than 100 by noting that 10 precincts “contained special voting booths where by law out-of-precinct voters can cast a ballot.” Although such precincts do exist, none are in these data set — a fact which can be verified in data made publicly available by the Mexican government (http://j.mp/mxife).7

Ultimately, we confirmed that our conclusions about coding errors held more generally with a formal Mexican Freedom of Information Act request and multiple conversations with officials at the Dirección Ejecutiva de Organización Electoral (DEOE) and Dirección Ejecutiva del Registro Federal de Electores at Mexico’s National Electoral Institute (INE). More detailed information about this issue appears in the Appendix to this paper and our Supplementary Appendix.

Unconventional Model Specifications

Finally, we consider the unusual model specifications in De La O (2013, 2015). Across electoral studies in American and comparative politics, researchers almost always measure the vote for an incumbent party as the incumbent vote share — the number of ballots cast for the incumbent party divided by the total number of ballots cast. Researchers from most countries then typically model vote share by including a lagged value of vote share as a control variable.

In contrast, De La O (2013, 2015) measured the incumbent vote for a party as the number of voters for that party divided by the voting age population — including voters, nonvoters, noncitizens, those ineligible to vote for other reasons, etc. The article then includes a lagged control variable which was not the incumbent vote share, and not the voting age population, but instead the number of people in the total population — including the voting aged population as well as all those under 18. As a result of this unusual decision, the outcome variable in this analysis does not add to 100% across parties, or across parties plus nonvoters, and the outcome and control variables are not logically or necessarily related. Although the model is theoretically possible — more noncitizens or infants in an area could in principle lead to more incumbent voting — this idea was not

7To be more specific, of the four types of polling stations in Mexico — básicas, contiguas, extraordinarias, and especiales — only especiales allow out-of-precinct voters. Yet, none of the polling stations included in the sample in De La O (2013) are of this type.
suggested in De La O (2013, 2015); indeed, to our knowledge, no other published statistical model of the electoral politics in any country and election has ever made these assumptions.

More importantly, because counts, unlike percentages, are unbounded, outliers on this scale greatly exacerbate the bias that results from influential outliers. This is why, as we will show, either correcting the coding errors or switching to the dominant method of constructing vote variables eliminates any support for the effects of programmatic policies on partisan outcomes.

Reanalyses

We begin our reanalysis with Figure 5 which replicates De La O (2013)’s least squares regression estimate and 95% confidence intervals for the intent-to-treat causal effect of Progresa on turnout (left panel) and incumbent vote (right panel). It does this for the data as coded in the original article and book (dotted vertical line) and for data from the official results among registered voters (solid vertical line). The horizontal line marking no effect appears at zero. The only point in including the outcome variables as originally measured (the dotted lines) is to complete our forensic data analysis and to show how correcting the statistical analysis method makes the result vanish; the large number of errors, including the two huge outliers we uncovered, indicate that the only valid causal estimates of the programmatic incumbent support hypothesis comes from the official vote and turnout figures (solid lines).

Reading each panel in Figure 5 from left to right, the different specifications we tried include the linear regression in the original article and book; a simple difference-in-means; a matching estimator; a regression controlling for log of population; a regression with lag turnout on the same scale as the outcome; and a regression after removing the two observations with the highest leverage. The panel on the right repeats all the analyses for incumbent (PRI) vote share.

For matching, we did Coarsened Exact Matching (CEM), adjusting the coarsening to deal with the presence of high leverage observations among the pre-treatment covariate. The distributions of covariates before and after matching are reported in Figures 7 and 8 in the Supplementary Appendix, with full information in our replication data set.
Figure 5: Intent-to-treat Causal Effects of Progresa on Turnout and Incumbent Party Vote. The left panel reports point estimates and 95% confidence intervals for the causal effect of Progresa on turnout as originally, and incorrectly, measured in De La O (2013) (dotted vertical lines), and for official turnout among registered voters (solid vertical lines) for several different specifications, including in De La O (2013) for the first pair on the left. The right panel repeats the same analyses for incumbent (PRI) vote share. Every estimate is indistinguishable from zero, except when using the wrong measure without controls (first two dotted lines in the right panel).

The results in Figure 5 exactly replicate results in De La O (2013, 2015) that give positive point estimates for turnout and vote share for the incumbent party, and a 95% confidence interval that excludes 0 for vote share but is insignificant for turnout. Using the original variable (with errors as is) reveals the same basic results, even using a simple difference in means estimator. However, once we use any of the four methods that each seek to control for the large imbalance induced by the data errors in De La O (2013), the effects vanish, with no statistically significant evidence for the effect of Progresa on either turnout or vote share. Moreover, rerunning any of the six analyses, while dropping the original error ridden variable and switching to official registration data, reveals no evidence of an effect for this nonpartisan programmatic policy on either partisan outcome, regardless of how the data are analyzed. Moreover, with the clean registration data, the confidence intervals are much narrower, and all twelve include zero as a causal effect.

Figure 6 repeats the same analyses, with the same robustness checks, for De La O
(2013, 2015)’s instrumental variable analysis estimate of the causal effects. The results here tell essentially the same story, indicating no statistically significant effect of non-partisan programmatic policies on voter turnout or vote for incumbent parties. Although again, only the official turnout and vote figures (solid lines) offer valid causal estimates.

Figure 6: Instrumental Variable Estimates of Progresa on Turnout and Incumbent Party Vote. In a manner directly parallel to Figure 5, this figure replicates the instrumental variable estimation from De La O (2013). Every estimate is indistinguishable from zero, except when using the wrong measure without controls (first two dashed lines in the right panel).

Many other alternative specifications and analyses of the results from De La O (2013) appear in our Supplementary Appendix, all leading to the same conclusion: Progresa had little or no effect on either voter turnout or the incumbent vote.

4 Generalizability and Future Research

We now address the question of how generalizable the consistent results from our two experiments may be, from an empirical and then theoretical perspective. This enables us to explore how our results fit into the literature and how future research might proceed on these and related questions.
Empirical Context  The policies studied in our two experiments have in common the four criteria listed in Section 1 — broad support in the legislature, objective rules for implementation, no exclusive credit claimed for services provision by any one party, and voter expectation of nonpartisan service provision. None of the observational studies that address, and sometimes support, aspects of the programmatic incumbent support hypothesis, analyze policies that fit all four criteria (Manacorda, Miguel and Vigorito, 2011; Pop-Eleches and Pop-Eleches, 2012; Zucco, 2013; Larreguy, Marshall and Trucco, 2015). The studies also analyze different policies and political situations than for the two experiments we discussed here. As such, the experimental and observational evidence on the question are, to some extent, addressing different questions. This suggests that the political context in which a policy is passed and implemented may be crucial.

For example, an important difference between the policies we examine and those analyzed in previous studies is the ability of incumbents to claim credit for the policies enacted. Consider that PANES, the temporary emergency relief program analyzed in Manacorda, Miguel and Vigorito (2011), was a one-time effort. As such, and as discussed by the authors, voters may have reasonably associated the policy with the party responsible for its enactment and implementation. Similarly, in the case of Romania, Pop-Eleches and Pop-Eleches (2012) explain that the Euro 200 program had a well known partisan intent. Finally, Larreguy, Marshall and Trucco (2015) study an interesting urban titling program crafted in 1973 when Mexico was not a democracy. Under this program, incumbents organized events where they claimed exclusive credit for the number of land titles granted during a given time period.

Similarly, in a field experiment in Uganda, Blattman, Emeriau and Fiala (2016) find that a program transferring cash to groups of unemployed youth increased support for and work on behalf of the opposition, among survey respondents. The policy was largely funded by the World Bank, and a large proportion of survey respondents credited this institution for the creation of the program instead of the country’s autocratic government. The authors hypothesize that this aspect of the program freed voters from clientelistic ties, thereby increasing their propensity to support and work for the opposition. In the
Phillippines, Labonne (2013) finds, based on a randomized evaluation, that a conditional cash transfer (CCT) program modeled after Brazil’s Bolsa Familia and Mexico’s Progresa increased support for local incumbents only in competitive (i.e., non-dynastic) municipalities with small federal budgetary transfers. The authors did not find similar electoral benefits in competitive municipalities with large transfers. The author’s explanation of this difference is that in competitive municipalities with large transfers, mayors distribute funds among individuals not receiving the CCT, making the voting behavior of program beneficiaries and non-beneficiaries indistinguishable from each other.

Together, these studies suggest the importance of context and policy design. They also suggest the considerable value of following up these studies so we can map out the details of the political and policy context and how they may affect electoral outcomes.

**Theoretical Evidence** A well developed formal theory literature in this area is almost solely concerned with the electoral effects of discretionary, rather than programmatic, spending policies. For example, Dixit and Londregan (1996, p.1132–1133) devote the first two pages of their article to clarifying this point. Theoretical results suggest that targeting tactical or pork barrel spending occurs most often to benefit swing voters (Dixit and Londregan, 1996) and marginal constituencies (Weingast, Shepsle and Johnsen, 1981), and this is indeed what the evidence shows (see Dahlberg and Johansson 2002 on the former and Primo and Snyder 2010; Keefer and Khemani 2009 on the latter). Studies have also found that such spending has substantial electoral payoffs (Levitt and Snyder, 1997; Evans, 2006).

One clever exception, proposed in De La O (2015), holds that minority parties will rationally choose to support a programmatic policy that will hurt them electorally if opposing it imposes even larger costs. We outline this theory here and then draw two new conclusions from it that suggest topics for future research.

In the version of this theory closest to the contexts in which Progresa and SPS were passed and implemented, the incumbent president’s party (P) proposes a CCT that is partisan (PCCT, i.e., over which incumbents have discretion) or nonpartisan (NCCT, i.e., a programmatic policy), and the median opposition party legislator (LO) then decides
whether to pass the program. Under the status quo, $P$ receives payoff $p^{sq}$. Denote the value of clientelism, which is realized only if PCCT is passed, as $v_1$ for $P$ and $v_2$ for $L_O$, where $v_1 > v_2$. Then, if PCCT is passed, the total payoff for $P$ is $p + v_1 - v_2$, where $p$ is the probability $P$ being reelected if NCCT is passed. The payoff for $L_O$ passing PCCT mirrors that of the incumbent: $1 - p - (v_1 - v_2)$. If $L_O$ rejects PCCT, the payoffs for $P$ and $L_O$ are $p^{sq} - e$ and $1 - p^{sq} + e$, respectively, where $e > 0$ “represents the payoff from supporting a nonclientalist poverty relief program, or rejecting a clientist program, when the other player does not” De La O (2015, p. 50). The book also assumes $p > p^{sq}$ “because the president can claim credit for the policy innovation.” Thus, $L_O$ will only pass PCCT if $1 - p - (v_1 - v_2) > p^{sq} - e$, a condition which never holds.

Alternatively, if $P$ proposes an NCCT, and the opposition passes it, $P$ and $L_O$ obtain payoffs of $p$ and $1 - p$, respectively. If the opposition does not pass the NCCT, the payoffs are $p^{sq} + e$ and $1 - p^{sq} - e$. Thus, De La O (2015) shows if $P$ proposes an NCCT, the opposition will pass it if the cost of passing the legislation is less than the cost of blocking it or, in other words:

\[ p - p^{sq} < e. \] (1)

Then, the incumbent, knowing that the opposition party will never pass a PCCT, is better off with the status quo than PCCT (since $p^{sq} > p^{sq} - e$) and so never proposes a PCCT in the first place. Then, if an NCCT that $P$ proposes doesn’t pass, $P$ gets payoff $p^{sq} + e$; if it passes, $P$ gets $p$. Thus, because $P$ is better off proposing an NCCT, regardless of what $L_O$ does, the equilibrium result is for $P$ to propose and $L_O$ to pass the NCCT.

The result in equation 1 from De La O (2015) shows that there exist conditions under which the opposition may pass policies that could hurt them. From this result, we now derive two new implications that may be worthy of further study.

First, we show that the theory is consistent with the opposite result, that the programmatic incumbent support hypothesis is false. Suppose that incumbents receive little or no benefits from passing an NCCT, i.e., $p \approx p^{sq}$. In this situation, Equation 1 still holds if $e$ is sufficiently large. As such, the theory is consistent with the programmatic incumbent support hypothesis being false and also with it being true; as a result, even if the theory
itself is true, it provides no information about the veracity of the programmatic incumbent support hypothesis.

Second, although the theory does not imply the programmatic incumbent support hypothesis, we analyze here whether it is possible to estimate the parameters of the theory (e.g., \( p, p^q \), and \( e \)) to test this hypothesis, as claimed by De La O (2015). As it turns out, this is not possible. The reason is that estimation would require observing the counterfactual case when an NCCT (or a programmatic policy more generally), was proposed but not passed. However, in the treated and control conditions of both experiments, the policy was proposed and passed for every observation, making it impossible to identify most parameters of the theory, including the causal effect of the opposition rejecting vs accepting the proposal, from either the SPS or Progresa experiment. Instead, each of the two experiments compares areas where the proposed-and-passed policy was implemented vs not implemented. This is an important quantity, relevant to the programmatic incumbent support hypothesis, but it cannot be used to test the theory in De La O (2015).

A valuable area for future research, then, would involve deriving formal theories that could shed light on when and under which political contexts and policy proposals the programmatic incumbent support hypothesis may or may not hold. It would also be useful if some of these theories could be directly tested from available data, such as cross-country data, and if they were tuned to the diverse areas in which experimental and observational data has been or could be analyzed.

5 Concluding Remarks

When incumbent parties exercise discretion over government spending to benefit specific groups of voters, an impressive literature shows that these voters reward the incumbents with electoral support in subsequent elections. The reason for this powerful result would seem to be called into question by claims that nonpartisan programmatic policies — policies over which incumbents have no discretion — have similar effects on voters making partisan decisions. We sought to contribute to an explanation of why this would happen and, also, why minority party politicians would, in the first place, support policies that
would lead voters to oppose them in the next election. It is perhaps fortunate, then, that our results are so strong in concluding that programmatic policies have little or no impact on voter turnout or voter support.

We are in the unusual position for our discipline of being able to test an important hypothesis with two extremely large scale, high quality randomized experiments. The policies evaluated in both cases were also of such large scale, and so highly impactful, it is unlikely that if an impact of programmatic policies on voter support does not exist in these cases that it could be found in other circumstances meeting the same four criteria as describe SPS and Progresa.

A Merging Census and Electoral Data

In many democracies, election administration is conducted in an entirely separate office than census operations, vital registration, and other demographic accounting. The result is that the definition of the areal units used in any data analysis that define electoral precincts often overlaps or conflicts with that for census geography. This causes common and well-known data issues, and must be treated carefully by any scholar using aggregate electoral data from many countries around the world (e.g., King and Palmquist, 1998).

A.1 The SPS Experiment

For the SPS experiment, we deal with the issues involved in having separate sources for electoral and census geographies in two ways. The first involves defining the “precinct cluster” as a new geographic entity for our unit of analysis. The second is our large scale, individual-level survey, for which no merging issues arise in the first place. These two data sources are the basis for the analyses in Figures 1 and 2, respectively.

We begin with available information, which includes, in addition to the electoral and census databases, (a) the set of villages that fall within (and help define) each health cluster according to the Health Ministry, (b) the set of villages that fall within (and help define) each electoral precinct according to the Electoral Institute, (c) the complete GIS definition of the precinct boundaries, (d) the geographic centroid for each village, and (e)
detailed satellite imagery.

We define the precinct cluster as the set of electoral precincts that contain at least one village belonging to a single health cluster assigned either treatment or control (but not both) within the SPS experiment. We do not use the textual name given to villages since they have different meanings in the two administrative databases.

Figure 7 gives examples of how we define precinct clusters in rural areas. In the left panel, two contiguous precincts (green areas with gray boarders) from the municipality of Ixtlahuaca each contain one village centroid (red dots). The precinct cluster in this panel is then the aggregation of both precincts (the entire green area). We also portray, in the right panel, the precinct cluster from the municipality of Santo Tomás; this precinct cluster, with three constituent precincts, is defined in the same way, even though it includes some villages (black dots) not participating in the experiment. We keep these precincts in the analysis, recognizing that including them could slightly attenuate the estimated effect of SPS on electoral outcomes in Figure 1 (but not Figure 2). Finally, we of course remove any precincts that have village centroids that span health clusters assigned to different treatment regimes.

Finally, creating precinct clusters in urban settings is slightly complicated by the fact that the urban health clusters are defined as aggregations of census tracts, and the tracts’ boundaries sometimes overlap precinct boundaries. We overcome this problem with detailed satellite imagery, which we used to check the population distribution across precincts and tracts. This allows us high levels of confidence that the overwhelming majority of the population in the precincts belongs exclusively to the health clusters, and our corresponding precinct clusters, that participated in the evaluation. This also eliminates any potential attenuation bias.

Figure 8 offers an example of the criteria we use to define precinct clusters in urban clusters. The figure displays a satellite image (with more resolution than we can print on the page) of a health cluster in Morelos in the experiment (red boundary) and the precincts, numbered 660 and 661 (blue boundaries), which it overlaps. We assigned a precinct to a health cluster if the population residing in the precinct is found almost exclusively within
Figure 7: **Defining Precinct Clusters in Rural Areas.** The figure shows how we create precinct clusters, with examples from the municipality of Ixtlahuaca (left panel) and the municipality of Santo Tomás (right panel). In each appears individual precincts (geographically contiguous green areas outlined in gray), village centroids from a health cluster randomly assigned to the control group (red dots), and a precinct cluster (the entire green area). On the right also includes a few village centroids from health clusters that were not part of the SPS experiment.

the area of the health cluster, as is the case in this example. In this particular case, the population reported by census officials in 2005 within the health cluster is 2,996 inhabitants. This figure is very close to the total population of 3,051 inhabitants that census and electoral officials report for precincts 660 and 661 during the same year. Therefore, we include these two precincts in our analysis and use them to define one precinct cluster.

**A.2 The Progresa Experiment**

**Name matching** The coding errors in De La O (2013) were generated by incorrectly using the textual name given to a village to match electoral and census data. Unfortunately, these names were never normalized or disambiguated, and the data contain no unique identifiers or matching keys. They are, in fact, created by a person in each office choosing or making up a name and labeling a geographic area, without coordinating with their counterpart in the other office. The result of this process is that different government
Figure 8: **Defining Precinct Clusters in Urban Areas.** The satellite image in this figure portrays precincts 660 and 661 (blue boundaries) and the overlapping health cluster in our experiment (red boundary). Because virtually all the population in the two precincts also fall within the health cluster, we assign a precinct cluster to be coincident with the health cluster.

Agencies often wind up using different names to refer to the same village or the same name to refer to different villages.

Figure 9 illustrates these errors with the two largest outliers from Figure 3. The goal of this analysis is to locate each village (the geographic centroid of which is portrayed as a dot) within the correct electoral precinct (the aerial unit in green). The left two panels follow the incorrect approach in De La O (2013) — assuming that electoral and census officials use identical village names (and distribution of the number of villages per precinct) to refer to the same geographic areas. This “name matching” procedure leads to the inclusion of the village of Ciudad Valles in precinct 266 of San Luis Potosí (top left panel) and Tulancingo in precinct 1502 of Hidalgo (bottom left panel). We now show that these matches are incorrect.
Figure 9: **Name Matching v. Correct GIS Location.** For precincts 266, San Luis Potosí (top row) and 1502, Hidalgo (bottom row), the left column assumes, as De La O (2013, 2015) that census and electoral officials rely on the same village names and numbers of villages per precinct and uses name-matching to locate villages. The right column uses accurate GIS coordinates of census villages.

**Accurate GIS Locations for the Progresa Experiment** To avoid entirely the problem De La O (2013) induced by name matching, we obtained from the Census Bureau the exact village centroids and mapped them with geographic information systems (GIS) technology into the known precinct boundaries (the two right panels in Figure 9). The errors can be seen clearly by the true locations of villages Ciudad Valles and Tulancingo (red dots) entirely outside the precincts (green regions), and instead in areas of high population density (as reflected by the large number of small-area precincts surrounding their
respective centroids).

The mistake leading to the errors in De La O (2013, 2015) are not the only two coding errors in the data. The article and book also incorrectly assumed that the number of included villages in each precinct by the electoral office was identical to that reported by the census office. For example, the electoral records indicate that precincts 266 and 1502 have 2 and 5 villages respectively. However, because of the population, disambiguation, and name matching problems, this count does not imply that they have 2 and 5 villages according to census records. In fact, the correct numbers, according to the precise GIS coordinates, are 6 and 10 villages, respectively.

These two errors turn out to be consequential mistakes. For example, the correct populations of the tiny villages in precinct 266 in 2000 are 4 (people), 82, 1, 2, 1, and 7. Yet, the village incorrectly included in this precinct had over 100,000 inhabitants. Similarly, the villages in precinct 1502 had populations in 2000 of 210, 205, 163, 156, 1998, 83, 97, 19, 41, and 3, whereas the village incorrectly included had 87,458 inhabitants. The same types of errors exist throughout the data in De La O (2013, 2015).

References


King, Gary, Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Clayton Nall, Ryan T.


