

Designing and Analyzing Randomized Experiments*

Yusaku Horiuchi[†]

Asia Pacific School of Economics and Government,
The Australian National University

Kosuke Imai[‡]

Department of Politics, Princeton University

Naoko Taniguchi[§]

Department of Sociology, Teikyō University

First Draft: March 8, 2005

This Draft: June 16, 2005

*We thank Larry Bartels, Rachel Gibson, Daniel Ho, Gary King, Matthew McCubbins, James Morrow, Alison Post, Jas Sekhon, Elizabeth Stuart, and seminar participants at the Australian National University, Harvard University, Princeton University, the University of California, San Diego, and the University of Michigan for their helpful comments. We are also grateful to Takahiko Nagano and Fumi Kawashima of Nikkei Research for administering our experiment, and Jennifer Oh and Claudia Ornelas for research assistance. We acknowledge financial support from the Telecommunications Advancement Foundation (Denki Tsūshin Fukyū Zaidan) and the Committee on Research in the Humanities and Social Sciences at Princeton University. An earlier draft of this paper was presented at the 2005 Annual Meeting of the Midwest Political Science Association, Chicago, April 7–10, 2005

[†]Lecturer, Asia Pacific School of Economics and Government, The Australian National University, Canberra ACT 0200, Australia. Phone: +61-2-6125-4295, Fax: +61-2-6125-5570, Email: yusaku.horiuchi@anu.edu.au.

[‡]Assistant Professor, Department of Politics, Princeton University, Princeton NJ 08544, U.S.A. Phone: +1-609-258-6610, Fax: +1-973-556-1929, Email: kimai@Princeton.Edu, URL: www.princeton.edu/~kimai.

[§]Assistant Professor, Department of Sociology, Teikyō University, 359 Ōtsuka, Hachioji, Tōkyō 192-0395, Japan. Phone: +81-426-78-3593, Email: n-oni@yb3.so-net.ne.jp.

Abstract

In this paper, we demonstrate how to effectively design and analyze randomized experiments, which are becoming increasingly common in political science research. Randomized experiments provide researchers with an opportunity to obtain unbiased estimates of causal effects because the randomization of treatment guarantees that the treatment and control groups are on average equal in both observed and unobserved characteristics. Even in randomized experiments, however, complications can arise. In political science experiments, researchers often cannot force subjects to comply with treatment assignment or to provide the information necessary for the estimation of causal effects. Building on the recent statistical literature, we show how to make statistical adjustments for these noncompliance and nonresponse problems when analyzing randomized experiments. We also demonstrate how to design randomized experiments so that the potential impact of such complications is minimized.

1 Introduction

In this paper, we demonstrate how to effectively design and analyze randomized experiments, which are becoming increasingly common in political science research (see e.g., Kinder and Palfrey, 1993; McDermott, 2002, for review). Indeed, the number of articles in major political science journals that analyze randomized experiments has more than doubled during the last decade.¹ Randomized experiments provide researchers with an opportunity to ascertain causal effects without bias. This is possible because the randomization of treatment makes the treatment and control groups equal on average in terms of both observed and unobserved characteristics.²

Even in randomized experiments, however, complications can arise (Barnard, Frangakis, Hill, and Rubin, 2003; Imai, 2005). In political science experiments, researchers often cannot force subjects to comply with treatment assignment or to provide the information necessary for the estimation of causal effects.³ Since experimental subjects can choose not to comply and/or not to respond, ignoring this selection process leads to invalid and inefficient causal inferences. Building on the recent statistical literature, we show how to make statistical adjustments for noncompliance and nonresponse when analyzing data from randomized experiments. We also illustrate how to design randomized experiments in ways that minimize the potential impact of such complications.

This paper offers four general methodological recommendations for researchers who design and analyze randomized experiments. First, *researchers should obtain information about back-*

¹We have examined all the articles that were published during the last decade in *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics*. From 1994 to 1999, only 11 articles (eight non-laboratory experiments) that use randomized experiments were published in these journals. However, this number increased to 25 (15 non-laboratory experiments) during the next five years (from 2000 to 2004).

²Randomization also guarantees that the treatment is causally prior to the outcome (i.e., no post-treatment bias and no simultaneity bias).

³While many studies simply ignore these problems, the studies that explicitly discuss the issues of noncompliance and/or nonresponse include Clinton and Lapinski (2004); Iyengar and Jackman (2003) and Mutz (2002).

ground characteristics of experimental subjects that can be used to predict their noncompliance, nonresponse, and the outcome. With such information, researchers can increase the validity of modeling assumptions that are necessary to obtain unbiased estimates of causal effects. Second, *researchers should conduct efficient randomization of treatments by using, for example, randomized block and matched pair designs.*⁴ These randomization schemes allow for more efficient estimation than would be possible under the simple randomization design, especially when the sample size is small. Third, *researchers must make every effort to record the precise treatment status of each experimental subject.* This information allows for greater accuracy in the interpretation of estimated causal effects as well as sensitivity analyses about modeling assumptions. Whenever possible, researchers should also check whether the administered treatment corresponds to the concept that one is trying to evaluate. Finally, *a valid statistical analysis of randomized experiments must properly account for both noncompliance and nonresponse problems.* We show that the Bayesian framework of Imbens and Rubin (1997) provides a flexible way to model such complications. We also offer easy-to-use software that implements the statistical model used in this paper.

To illustrate our methodology, we designed and conducted an Internet survey-based randomized experiment during Japan’s 2004 Upper House election where randomly selected eligible voters were encouraged to view the specific policy information available at the designated official party websites. We analyzed this experimental dataset to examine whether more information leads to a higher voter turnout, as many political scientists have hypothesized. Our experiment is of particular interest for two reasons. First, major political parties in Japan have recently begun to prepare “manifestos” that explicitly state formal proposals for major policy issues. Second, the party websites are becoming an increasingly important medium through which Japanese voters access policy information. Our results indicate that voters are less likely to abstain when they

⁴Randomized block designs are implemented and discussed later in this paper. Matched-pair designs refer to a randomization scheme where experimental subjects are paired based on their background characteristics. The complete randomization of treatment is then conducted within each pair.

were exposed to the policy information on the websites of both *both* ruling and opposition parties. We also find that the information effect is larger among those voters who were planning to vote but were undecided about which party they were going to vote for in the election.

The rest of the paper is organized as follows. In Section 2, we briefly discuss the existing literature on information and voter turnout as well as the advantages of randomized experiments for testing the information hypothesis. In Section 3, we present the design of our Internet survey-based randomized experiment and discuss general issues in designing randomized experiments. In Section 4, we describe our statistical methods that account for noncompliance and nonresponse in randomized experiments. The results of our statistical analysis are shown and discussed in Section 5. Section 6 concludes.

2 Information Hypothesis and Randomized Experiments

We illustrate our methodologies by testing one of the most fundamental hypotheses about voting behavior. Since the publication of Anthony Downs' (1957) influential book, political scientists have emphasized information as a key determinant of voting behavior (e.g., Alvarez, 1998; Ansolabehere and Iyengar, 1995; Ferejohn and Kuklinski, 1990; Grofman, 1993; Huckfeldt and Sprague, 1995; Iyengar and Kinder, 1987). In particular, Downs argued that the less information a voter has, the more likely he is to abstain. This information hypothesis inspired studies by later scholars and led to the development of various formal models that attempt to explain the causal mechanism of the information effect on turnout (e.g., Feddersen and Pesendorfer, 1999; Ghirardato and Katz, 2002; Matsusaka, 1995; Palfrey and Rosenthal, 1985). The assumptions and logic of these formal models are different, but their basic implications have been largely consistent with Downs' original conjecture.⁵

⁵We do not test the different causal mechanisms derived from these formal models. This limitation arises in part because the key causal variables used in many formal models are difficult, if not impossible, to manipulate in a real election (e.g., the number of voters of different types in an election).

Although randomized experiments have their own limitations, there is a significant advantage when estimating the causal effect of information on voter turnout. Namely, an experimental study allows us to manipulate the quantity and/or quality of additional information each respondent receives during an actual election. Survey researchers have designed various questions to measure the amount of information voters possess (e.g., Bartels, 1996; Lupia, 1994; Rosenstone and Hansen, 1993). Although the literature yielded considerable insight about the possible *association* between voting and information, it faces a common methodological challenge that hinders the estimation of the *causal effect* of information on voting behavior. The problem is that those voters who have a strong intention to vote may be more likely to acquire the information. Our experimental approach is designed to address this problem of *endogenous information acquisition* in real election settings (see also Iyengar and Jackman, 2003; Wantchekon, 2003). Conducting experiments is a powerful approach, but it is certainly not the only way to solve this problem. A careful analysis of observational data with appropriate assumptions can also shed light on causal relationships between information and voting behavior (e.g., Sekhon, 2004; Lassen, 2005).

An important feature of our experiment is the use of the Internet. In many democracies including Japan, the Internet has become an important source of information for voters and an essential tool for parties and candidates to reach voters (e.g., Dulio *et al.*, 1999; Gibson *et al.*, 2003a,b). Our study contributes to the emerging literature on the question of how the Internet affects voters' political attitudes, opinions, and voting behavior (e.g., Johnson and Kaye, 2003; Lupia and Philpot, Nd; Weber *et al.*, 2003). We also hope that our experimental design and analysis together serve as a template for researchers who are planning to design, conduct, and analyze Internet survey-based randomized experiments.⁶

Finally, Japan is an interesting case when examining the causal effects of policy information.

⁶For example, Time-Sharing Experiments for the Social Sciences (TESS), funded by the National Science Foundation, offers researchers opportunities to conduct Internet-based experiments. See <http://www.experimentcentral.org>.

In particular, the country recently underwent significant electoral and political reforms in order to minimize the influence of personal politics and to promote policy-based electoral campaigns and voter participation. As a part of this effort, major Japanese political parties have begun to prepare “manifestos” that explicitly state their formal policy proposals on major issues. Can these manifestos attract voters and revive the declining interest in politics? Does manifesto-based electioneering encourage voters to cast their ballots based on party policies rather than on personal connections to particular candidates? These questions are particularly important in the context of Japanese politics.

3 Designing Randomized Experiments

In this section, we describe our design of the randomized experiment that we conducted through a Japanese Internet survey firm, *Nikkei Research*, in June and July of 2004.⁷ This Japanese election experiment illustrates how to incorporate three out of our four methodological recommendations when designing randomized experiments; (1) to collect background characteristics that are key predictors of noncompliance, nonresponse, and turnout, (2) to conduct efficient and accurate randomization of treatments via randomized block designs, and (3) to record the treatment status as precisely as possible. Although experiments with different goals often require different designs, we believe that these methodological points are fairly general and can be incorporated into many experimental studies. Our experiment consisted of three separate surveys, as depicted in Figure 1—a screening survey, a pre-election survey, and a post-election survey. In the following, we explain the implementation of each survey and our experimental design in detail.

⁷The company has a sampling pool of roughly 40,000 Internet users throughout Japan who have agreed to receive occasional electronic mails asking them to participate in online surveys. Those who fill out a survey questionnaire have a chance to win a gift certificate in the amount of approximately five to ten dollars.

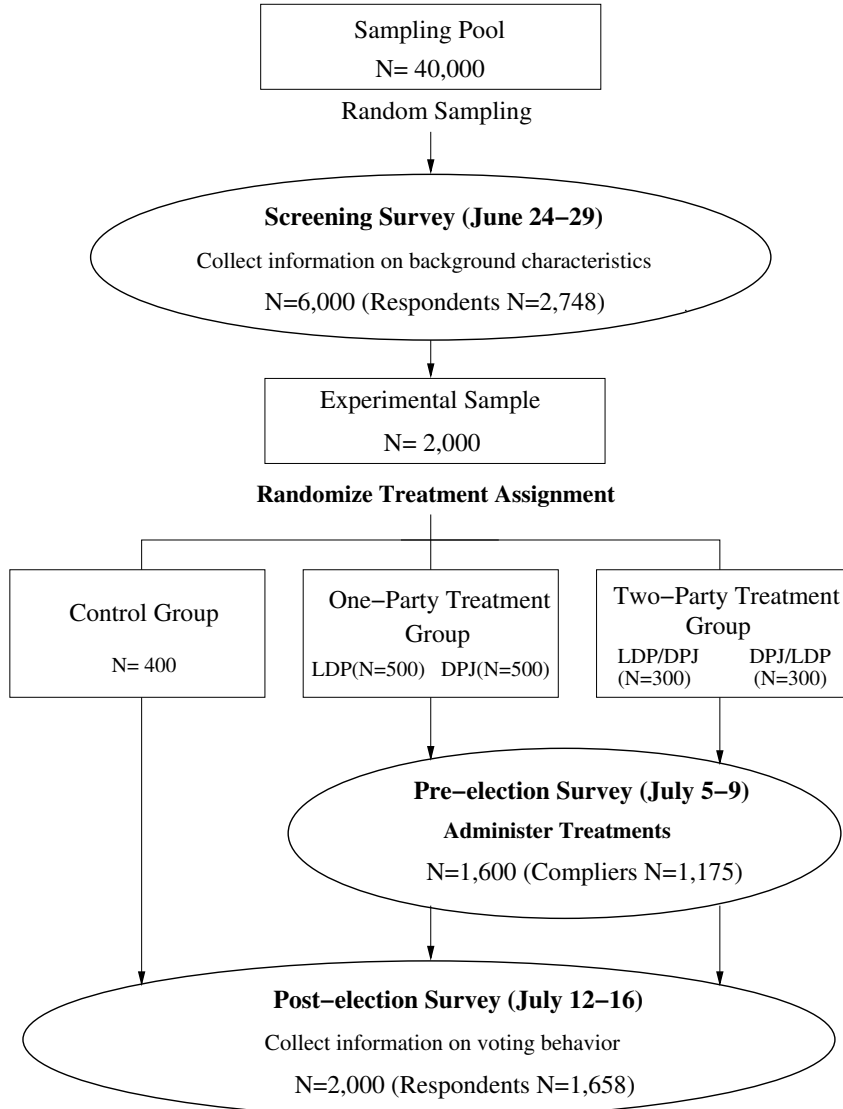


Figure 1: The Design of the Japanese Election Experiment. The experiment uses three surveys. The screening survey measures the pre-treatment covariates, and the pre-election survey administers the randomized treatments. Finally, the post-election survey measures the outcome variable. The election was held on July 11.

3.1 Screening Survey

Between June 24 and 29, two weeks prior to the election day (July 11), we conducted a screening survey and asked 6,000 randomly sampled respondents to answer several questions about them-

selves and their voting intention in the upcoming election.⁸ The main purpose of the screening survey was to collect the background characteristics of experimental subjects, which are important predictors of noncompliance, nonresponse, and the outcome. Therefore, the first part of the survey questionnaire asked for each respondent's prefecture of residence, age, gender, and highest education completed, which are known to be key demographic variables for Japanese voting behavior. We also asked the respondents about their party preference,⁹ and whether they were planning to vote in the upcoming election (planning to vote, not planning to vote, or undecided) and, if so, which party and/or candidate they were going to vote for, and how much confidence they had in their voting plan (a four-point scale).¹⁰ We measured these variables because they are powerful predictors of the outcome variable, i.e., voter turnout. To avoid further statistical assumptions at the data analysis stage, the survey was designed so that respondents had to answer all of the questions in order to complete the survey.

Of 6,000 individuals who received an electronic mail asking them to fill out the screening survey, 2,748 individuals completed the survey. From this group, we then randomly selected 2,000 eligible voters as our experimental sample.¹¹ With a few exceptions of over-represented urban prefectures, a comparison of the 2000 Census data and our sample shows no clear evidence of geographical sampling bias. Yet, as is the case for many other experimental studies, our sample is not fully representative of the Japanese electorate. First, the individuals in our sample were those who

⁸We asked *Nikkei Research* to select randomly an equal number of male and female Internet users, all of whom were between ages 20 and 59. Note that Japan, unlike the United States, automatically registers every eligible voter (20 years old and over). Furthermore, the wording of our survey questions closely followed the standard protocol from large-scale Japanese national surveys, such as those from the Japan Election Study (JES) projects.

⁹For party preference, we first asked respondents to indicate whether they supported a particular party or a candidate, and then asked them to rate their support for each of the five major parties on a four-point scale.

¹⁰Both the Upper and Lower House elections in Japan adopt a combination of plurality and proportional representation systems, with each voter casting two ballots.

¹¹The sample size is reduced to 2,000 given the financial constraint.

had Internet access, voluntarily registered themselves for the survey firm, and agreed to fill out the screening survey. The lack of representativeness and the unavailability of sampling weights are clear weaknesses of the Japanese Internet survey company used in our experiment. We note, however, that some Internet survey firms in other countries may provide researchers with a more representative sample and/or sampling weights.¹²

A lack of representativeness also means that our sample is likely to contain active Internet users with some interest in politics. This may not necessarily be a negative factor if the research goal is to examine the effectiveness of a party website among Internet users. Another consequence is that, as seen later, the rates of noncompliance and nonresponse are much lower than typical randomized experiments (see Imai, 2005). Although one ideally wants a representative sample of a target population, the high response and compliance rates bring their own advantages; the results rely less on statistical assumptions than would be possible with a representative sample. These issues reflect a usual trade-off between internal validity (i.e., validity of statistical assumptions) and external validity (i.e., representativeness of sample) in empirical research.¹³

3.2 Pre-election Survey

From July 5 to 9, we conducted the pre-election survey. The main purpose of this survey was to administer the treatments. The Upper House election was held on July 11, two days after the closing of the survey. Before sending an electronic mail soliciting their participation in the survey, we randomly assigned the treatments to the voters of our experimental sample. In particular, we considered two types of treatments and randomly divided the sample into three groups, i.e., the two treatment groups and the control group. The voters in the *one-party treatment group* were asked to visit the designated website of either the Liberal Democratic Party (LDP) or the

¹²For example, scholars of U.S. politics use Knowledge Networks (<http://www.knowledgenetworks.com>), which provides such information.

¹³Experimental studies typically gain internal validity at the cost of external validity of observational studies.

	Randomized blocks						Total
	I	II	III	IV	V	VI	
	Planning to vote Male	Female	Not planning to vote Male	Female	Undecided Male	Female	
One-party treatment group							
DPJ website	194	151	24	33	36	62	500
LDP website	194	151	24	33	36	62	500
Two-party treatment group							
DPJ/LDP websites	117	91	15	20	20	37	300
LDP/DPJ websites	117	91	15	20	20	37	300
Control group							
no website	156	121	19	26	29	49	400
Block size	778	605	97	132	141	247	2000

Table 1: Randomized Block Design of the Japanese Election Experiment: Six randomized blocks were formed on the basis of the two covariates, gender (male or female) and the answer to the question, “Are you going to vote in the upcoming election?” (“planning to vote”, “not planning to vote”, or “undecided”). Within each block, the complete random assignment of the treatments is conducted so that the size of each treatment and control group is equal to the predetermined number. The total sample size is 2,000.

Democratic Party of Japan (DPJ), while those in the *two-party treatment group* were asked to visit the websites of both parties.¹⁴ Our experimental design of two different treatment conditions allows us to examine not only whether exposure to policy information influenced voter turnout, but how different levels of exposure affected turnout. Finally, none of the individuals in the control group were asked to participate in the pre-election survey. In general, a control group is essential for causal inference that requires counterfactual analysis.¹⁵

In order to randomly divide the sample into two treatment and control groups, we applied the

¹⁴The LDP has been in power since its foundation in 1955 (except during a short period from 1993 to 1994), and the DPJ is the largest opposition party, formed by amalgamation of various parties. In this election, 82 percent of seats were won by these two major parties.

¹⁵This is especially true and consequential when the problem of noncompliance exists. In such a situation, a direct comparison of different treatment groups becomes difficult because compliers and noncompliers must be identified separately for each treatment. This problem often occurs in field experiments (Imai, 2005).

randomized block design shown in Table 1. We formed six blocks on the basis of the gender and voting intention variables, which we obtained from the screening survey. We chose these variables because they are expected to be important predictors of Japanese voters' turnout decision (e.g., Cox *et al.*, 1998; Pharr, 1981, pp.25–26). Within each of six randomized blocks, we conducted the complete randomization of treatments such that the total number of voters is 1,000, 600, and 400 for the one-party and two-party treatment groups, and the control group, respectively. Within the one-party treatment group, we randomly selected half of the voters and instructed them to visit the DPJ website. The other half was instructed to visit the LDP website. Similarly, for the two-party treatment group, a random half of the voters was instructed to visit the DPJ website first before visiting the LDP website, while the order was reversed for the other half.

An important advantage of the randomized block design is that it effectively reduces random as well as systematic differences (in a particular coordinate defined by selected pre-treatment covariates) between the treatment and control groups. The resulting estimates are more efficient than those obtained under a simple randomization design (e.g., Cox and Reid, 2000). Thus, it is recommended to select the covariates that are good predictors of the outcome variable when forming randomized blocks. Other types of experimental designs based on a similar idea are also possible. For example, one can use matched-pair designs by creating pairs (groups) of observations with similar characteristics (e.g., through a matching method based on the Mahalanobis distance measure) and conduct complete randomization within such pairs (groups) (e.g., Hill *et al.*, 1999; Greevy *et al.*, 2004).

Both randomized blocks and matched-pair designs aim to ensure accurate and efficient randomization of treatments. Table 4 in Appendix A illustrates the overall balance of the observed covariates. Indeed, none of the t -statistics is significant at conventional levels, showing that as expected, our randomized block design produced a good balance of observed covariates.¹⁶ There are only small differences between the treatment and control groups in terms of the observed

¹⁶We use t -statistics here, but more sophisticated measures of balance might be necessary in some situations.

covariates including age, education, party preferences, and so forth.

For both LDP and DPJ, we used the official website that shows their party platform. Specifically, we selected the particular section of the two parties' official "manifesto" that describes their formal policy proposals on pension reform.¹⁷ Our focus on pension reform was motivated by the fact that it was one of the two major issues in this election along with the government's policy towards Iraq.¹⁸ The LDP website presented their pension reform policies by explaining in detail the legislation that had been passed in the Diet approximately one month before the election. This legislation was widely considered to constitute a set of minor changes to the then current pension system. The LDP website gave little information about plans for further reforms, such as the widely-debated future integration of various pension systems. It also did not mention other unresolved and controversial issues.¹⁹ In contrast, at the very beginning of its party manifesto, the DPJ emphasized the need for national integration of pension systems and proposed the abolishment of the special pension for Diet members. While the DPJ website proposed major reforms, it did not specify the content of the reforms and did not explain how such proposals would be implemented.

Before being instructed to visit the website in the pre-election survey, voters were presented with a few brief questions about the pension reform. These were general questions that were

¹⁷In Japan, parties and candidates are not allowed to change the contents of their websites during a campaign period. This regulation is convenient for our experimental study, as we need not consider the possibility that different voters within the same treatment group viewed different web contents.

¹⁸According to a poll conducted by *Asahi Shimbun* and the University of Tokyo, 53 percent of candidates and 61 percent of voters regarded pension reform as one of the most important issues in this election (*Asahi Shimbun*, evening edition, June 24, 2004).

¹⁹These issues included the question of whether to abolish the special pension scheme available for Diet members, the political scandal about the Diet members who had not been paying for their national pension premiums, and the politically sensitive question of whether to raise the consumption tax to pay for the ever-increasing burden of the pension system.

intended to prepare voters before being exposed to the policy information. After answering these questions, voters were instructed to click a direct link which took them to the designated party website. The instruction also included a friendly warning, which mentioned they would be asked about their opinions on the website after visiting it. We designed the survey so that voters would have to visit the website in order to go to the next question. In addition, we also obtained information as to whether and how long voters actually opened the designated website in their browser even when the voters decided not to go to the next question. Knowing the treatment status of each subject as precise as possible is a very important consideration because the interpretation of estimated causal effects is difficult without such information. The goal of our experiment is to estimate the causal effect of being exposed to policy information through viewing a party website. However, if one wishes to know why policy information increases/decreases turnout, it is equally important to check whether the administered treatments actually correspond to the concepts of underlying theoretical models that one is trying to evaluate.²⁰

Finally, to complete the survey, voters were asked to answer several brief questions about the website they had just visited. For those voters who were assigned to the two party websites, the same set of questions was presented after they visited each website. At the end of the survey, voters were given a chance to write freely their opinions about the website. Although filling in this open-ended question was optional, nearly 80 percent of those who participated in the pre-election survey wrote some comments and/or opinions. The voters indicated a high level of interest in the pension reform, the upcoming election, and/or the party websites.

3.3 Post-election Survey

The day after the July 11 election, we started a post-election survey, which closed on July 16. The goal of this survey was to measure the outcome variable for all 2,000 experimental subjects. We

²⁰This can be often done by asking respondents additional survey questions. However, this may introduce extra complications and lead to higher nonresponse rate.

used the same questionnaire for everyone, asking whether they had voted in the election. We kept the survey short in order to minimize the unit nonresponse, and as a result, more than 80 percent of the respondents completed the survey. In the next section, we demonstrate how to adjust for this nonresponse problem statistically when estimating causal effects in randomized experiments with noncompliance.

4 Analyzing Randomized Experiments

In this section, we present the general statistical framework of Imbens and Rubin (1997) that can be used to analyze randomized experiments. As demonstrated by Barnard *et al.* (2003) and others, this framework provides flexible modeling strategies that account for noncompliance and nonresponse. These two problems are absent in the ideal prototype of classical randomized experiments, but nevertheless are common in non-laboratory experiments. Based on this framework, we define the causal effects of information on voter turnout. We then develop our statistical model and describe how to estimate the causal quantities of interest based on the model.

4.1 Randomized Experiments with Noncompliance and Nonresponse

In political science experiments, researchers often do not have full control over their human subjects and are likely to face additional complications that call for statistical adjustments. In particular, the problems of noncompliance and nonresponse frequently threaten the internal validity of randomized experiments, especially when they are conducted outside the laboratory. In the context of our Japanese election experiment, some voters did not visit the designated website even when they were instructed to do so (i.e., noncompliance). Moreover, some voters did not fill out the post-election survey, and therefore the outcome variables were not recoded for them (i.e., nonresponse). Since these two problems typically do not occur completely at random, ignoring them in estimation may severely bias causal inference.

To analyze randomized experiments with noncompliance and nonresponse, we begin by describing the formal statistical framework of Imbens and Rubin (1997). We use our Japanese election experiment as a concrete example to illustrate the applicability of this general statistical framework to the analysis of experimental data. We conduct three separate analyses with different binary treatment variables. Although it is possible to define a multi-value treatment and conduct appropriate analyses (see Imai and van Dyk, 2004), we analyze binary treatments for the sake of simplicity. First, we examine the causal effect of browsing one party website and two party websites, separately. In addition, we estimate the causal effects of visiting at least one designated party website for different subgroups in the sample that we select on the basis of observed pre-treatment covariates.

Formally, let Z_i be the treatment assignment indicator variable, which is equal to 1 if voter i is instructed to visit a party website (or party websites) and is equal to 0 otherwise. Next, use T_i to represent the actual treatment indicator variable, which is equal to 1 if voter i actually visits the website (or the websites) and is equal to 0 otherwise. If a voter logs on to the pre-election survey questionnaire website but does not visit the designated website, we set $T_i = 0$ for the voter. There were 133 such individuals, 63 of which belong to the one-party treatment group. This corresponds to about 10 percent of the voters who logged onto the survey website. In the two-party treatment group, there were 18 voters who visited only one designated website and did not complete the pre-election survey. We set $T_i = 0$ for these voters.²¹ Because of such drop-outs, we conduct sensitivity analyses using three different definitions of actual treatment status (logged on to the survey website, visited the party websites, and completed the survey). The detailed information about the treatment status we collect allows us to conduct such comprehensive sensitivity analyses.

We can now define two types of individuals in our experiment – compliers and noncompliers. Compliers refer to the voters who visit the designated party website *only when* instructed to do so (i.e., $(T_i = 1, Z_i = 1)$ and $(T_i = 0, Z_i = 0)$), while noncompliers are those who do not follow the

²¹Everyone who visited the two party websites completed the survey (i.e., no drop out after visiting both websites).

instructions. There are three types of noncompliers (Angrist *et al.*, 1996); *always-takers*, who do visit the party website regardless of whether they are instructed to do so (i.e., $(T_i = 1, Z_i = 1)$ and $(T_i = 1, Z_i = 0)$), *never-takers*, who do not visit the party website regardless of the instruction (i.e., $(T_i = 0, Z_i = 1)$ and $(T_i = 0, Z_i = 0)$), and *defiers*, who visit the website only when they are not instructed to do so (i.e., $(T_i = 0, Z_i = 1)$ and $(T_i = 1, Z_i = 0)$). We use C_i as the complier indicator variable, which is equal to 1 if respondent i is a complier and equal to 0 if he is a noncomplier.

In our analysis, we assume that there are neither always-takers nor defiers.²² Although these assumptions are not testable from the observed data,²³ they appear to be reasonable in our experiment. Only a small number of voters would have looked at the party website spontaneously. For example, one survey shows that only 3 percent of the respondents visited websites of parties and candidates during the 2003 Lower House election.²⁴ Moreover, it is also unlikely for ordinary voters to obtain an official manifesto. Article 142 (2) of the Public Office Election Law (Kōshoku Senkyo Hō) allows only limited distribution of manifestos by political parties. Indeed, another survey shows that even during the 2003 Lower House election, which was the first manifesto election in Japan, only 6 percent of the respondents obtained a complete version of a manifesto while 10 percent saw an outline of one.²⁵

Figure 2 shows that from the observed data, i.e., treatment assignment Z_i and actual treatment T_i , we can identify the compliance status of the voters in the treatment group (i.e., those in the upper and lower left cells of the figure). However, for the voters in the control group (i.e., those in

²²In the literature, the assumption of no defiers is called monotonicity.

²³We were unable to obtain whether the individuals in the control group accessed these specific party websites.

²⁴The survey is conducted by *Asahi Shimbun* and the University of Tokyo. <http://politics.j.u-tokyo.ac.jp/data/data01.html> (accessed on 12 June 2005).

²⁵The survey is conducted by Fuji Sōgo Kenkyūjyo in November 2003. http://www.mizuho-ir.co.jp/research/documents/manifesto031113_report.pdf (accessed on 12 June 2005).

		Treatment Assignment	
		$Z_i = 1$	$Z_i = 0$
Actual Treatment	$T_i = 1$	Complier $Y_i(1)$ is observed; $C_i = 1$	
	$T_i = 0$	Noncomplier $Y_i(0)$ is observed; $C_i = 0$	Complier or Noncomplier $Y_i(0)$ is observed; $C_i = ??$

Figure 2: Compliers and Noncompliers in the Japanese Election Experiment. The figure classifies compliers and noncompliers by treatment assignment, Z_i , and actual treatment, T_i . We assume that noncompliers solely consist of never-takers. From the observed data, Z_i and T_i , one can identify compliers and noncompliers in all but one case where $Z_i = T_i = 0$. The upper right cell is empty because we assume that always-takers and defiers do not exist in this experiment.

the lower right cell of the figure), we need to infer their compliance status using the observed compliance pattern of the treatment group. As we shall see, randomization of treatment assignment Z_i makes such inference possible because it guarantees that voters in the control group are similar to those in the treatment assignment group in terms of their observed and unobserved characteristics. In our experimental sample, the proportion of compliers is estimated to be about 70 percent (see Section 5), which is high when compared to typical field experiments in political science.²⁶ A high compliance rate is crucial for successful statistical analyses of randomized experiments because it reduces the degree to which estimated causal effects rely on modeling assumptions that are often difficult to verify from observed data.

4.2 Definition of Causal Effects

Following Rubin (1974) and Holland (1986), we define two potential outcomes, $Y_i(1) \equiv Y_i(T_i = 1)$ and $Y_i(0) \equiv Y_i(T_i = 0)$. We observe $Y_i(1)$ (i.e., a binary variable of turnout) for voter i if she visits the designated website, while $Y_i(0)$ is observed if she does not visit the website. This means we only observe one of the two potential outcomes, and the observed outcome variable can be

²⁶For example, the compliance rate of Gerber and Green’s (2000) experiment is as low as 25 percent (Imai, 2005).

defined as $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$.²⁷ Moreover, there exists the nonresponse problem because some respondents do not fill out the post-election survey.²⁸ To introduce a formal notation, let R_i represent an indicator variable which is equal to 1 if Y_i is observed and equal to 0 if it is missing. In our data, 342 respondents out of 2,000 individuals did not fill out the post-election survey (75 of them belonged to the control group, 113 of them were members of the one-party treatment group, and the others – 154 voters – belonged to the two-party treatment group).

Given this setup, we define a causal effect of information on turnout for voter i as the difference between the two potential outcomes: $Y_i(1) - Y_i(0)$. A fundamental problem of causal inference is that we can observe only one of two potential outcomes, but the calculation of a causal effect requires both of them (Holland, 1986). Furthermore, implicit in this formulation is the assumption of *no interference among units* (Cox, 1958; Rubin, 1990). That is, for all $i \neq j$, we assume that voter i 's treatment assignment status Z_i , does not affect voter j 's treatment status T_j (provided that Z_j is constant), and that the potential outcomes for voter j , $Y_j(1)$ and $Y_j(0)$, are not affected by the treatment status of voter i , T_i , or her treatment assignment status Z_i . In our experiment, this assumption seems reasonable because the voters in our sample are unlikely to communicate with each other about the experiment.²⁹

We also assume *no direct effect of treatment assignment*. That is, for voter i , the treatment assignment status Z_i is assumed to affect the voter's potential outcomes, $Y_i(1)$ and $Y_i(0)$, only

²⁷More generally, the potential outcomes are defined as a function of both actual treatment and treatment assignment i.e., $Y_i(T_i, Z_i)$ (see Angrist *et al.*, 1996). Under our assumptions, however, this is not necessary. Therefore, we avoid the general formulation for the sake of notational simplicity.

²⁸Note that our design of the screening survey guarantees that we do not have any missing values for the covariates.

²⁹We emphasize that this assumption is often implicitly invoked in both experimental and observational studies. In principle, one can relax the assumption by directly modeling the dependence between the potential outcomes of voter i and the treatment status of voter j . Doing so, however, significantly complicates the analysis and requires additional assumptions.

through the actual treatment status, T_i . This means that the causal effect is zero for never-takers because their treatment status is the same ($T_i = 0$) regardless of their treatment assignment status Z_i .³⁰ In our experiment, the assumption is violated if a respondent changes her decision to vote because she is instructed to visit a party website even though she does not actually complete the pre-election survey. This scenario is a potential concern for 113 voters who logged onto the pre-election survey but did not complete it. Therefore, as mentioned earlier, we conduct sensitivity analyses by applying various definitions of actual treatment status. This point directly relates to the third of our four methodological recommendations that we set forth in Section 1. The detailed information about the treatment status helps to clarify the interpretation of estimated causal effects and to evaluate the plausibility of statistical assumptions.

Our quantities of interest are defined separately for these two types of voters (see also Angrist *et al.*, 1996). By definition, noncompliers never receive treatment and so the effect of actual treatment T_i cannot be inferred from the observed data for this subgroup. Moreover, under our assumption, the effect of treatment assignment Z_i is zero for noncompliers. Therefore, we may focus on the causal effect of the treatment for compliers who are in the upper left and lower right cells of Figure 2. The estimand, the sample complier average causal effect or CACE is defined as,³¹

$$\text{CACE} = \frac{\sum_{i=1}^N C_i [Y_i(1) - Y_i(0)]}{\sum_{i=1}^N C_i}. \quad (1)$$

In our experiment, CACE defines the causal effect of information on voting behavior *for those who visit the website only when told to do so*. It is important to note that CACE does not equal the usual sample average treatment effect or ATE, $\frac{1}{N} \sum_{i=1}^N [Y_i(1) - Y_i(0)]$, which is the causal effect for the entire sample. From the perspective of policymakers who want to increase voter turnout by using the Internet, CACE might be of greater interest than ATE because political parties can

³⁰In the literature, this assumption is called exclusion restriction for never-takers. One can make a similar assumption about always-takers if they exist.

³¹Another estimand of interest is its population counterpart (see Imbens, 2004).

not force every voter to visit their party website. Hence, it is important to estimate the causal effect for those voters who are likely to be exposed to policy information via the Internet.

Another quantity of interest is the intention-to-treat effect or ITT effect, which represents the average causal effect of *treatment assignment*, Z_i , rather than the actual treatment, T_i .³² In our experiment, this is the causal effect of being asked to visit the party website rather than the effect of actually visiting the website. Unlike CACE, the ITT effect does not directly correspond to the effect of actual treatment. However, it represents the effectiveness of the treatment assignment for the whole sample, rather than for a subsample. The ITT is also typically easier to estimate in randomized experiments because it does not involve the identification of unobserved compliance status. Despite this important difference, CACE and the ITT effect are closely related because CACE represents the ITT effect for compliers (this follows from the definition of compliers). Furthermore, the ITT effect (for the whole sample) is the weighted average of ITT effects for compliers and noncompliers.³³ The relationship implies that a low compliance probability leads to a larger difference between CACE and the ITT effect. Since our assumption implies that the causal effect for noncompliers is zero, CACE will be always larger than the ITT effect.

Earlier in this section, we argued that always-takers are unlikely to exist in our experiment. Fortunately, we also know the direction of bias that arises even when ignoring their existence. While the estimation of ITT effect is not affected at all, the CACE will be *underestimated*. The size of this bias depends on the proportion of always-takers in the sample. This illustrates the danger of high noncompliance in randomized experiments.

³²If one adopts the general formulation of potential outcomes in footnote 27, then ITT effect can be defined as $\sum_{i=1}^N [Y_i(T_i, 1) - Y_i(T_i, 0)]/N$.

³³Under our assumptions, the ITT effect equals CACE multiplied by the fraction of compliers in the sample, $\sum_{i=1}^N C_i/N$.

4.3 Model, Estimation, and Inference

We now describe our model based on the assumptions explicitly stated above. Our model consists of two parts. First, we model the conditional probability of being a complier given each voter’s observed covariates. We use the following binary probit model with linear predictors,

$$\Pr(C_i = 1 \mid X_i, \xi) = \Phi(X_i^\top \xi), \quad (2)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, X_i represents the observed pre-treatment covariates for respondent i , which includes indicator variables for each of the randomized blocks,³⁴ and ξ is the vector of coefficients. If noncompliers include always-takers as well as never-takers, we model the compliance status as a multinomial outcome with three categories using, for example, multinomial logit or probit models (e.g., Hirano, Imbens, Rubin, and Zhou, 2000; Imai and van Dyk, 2005).

Second, we model voter turnout given the compliance status C_i and the treatment status T_i , as well as the observed covariates X_i . Again, we use a binary probit model,

$$\Pr(Y_i = 1 \mid C_i, T_i, X_i, \alpha, \beta, \gamma) = \Phi[\alpha C_i T_i + \beta C_i (1 - T_i) + X_i^\top \gamma], \quad (3)$$

where α and β are the intercepts specific to compliers with and without the treatment, respectively, and γ is a vector of coefficients for X_i . The base category is noncompliers. Depending on the type of the outcome variable, researchers can chose an appropriate model (e.g., a Normal linear model for a continuous outcome variable).

Note that the value of the outcome variable, Y_i , is missing for approximately 17% of the experimental subjects. However, deleting the observations with nonresponses, as often done in practice, results in bias and inefficiency of causal inferences (e.g., Frangakis and Rubin, 1999). The bias emerges because the listwise deletion is based on an optimistic assumption that the data are missing completely at random. The inefficiency problem arises from information loss due to the

³⁴It is also possible to treat block effects as random rather than as fixed. See for example Smith (1973).

exclusion of some observations from the analysis.³⁵ More importantly, such an analysis completely changes the qualitative interpretation of the estimated causal effects because the resulting inference is limited to those voters who participated in the post-election survey.

Therefore, we address the nonresponse problem by modeling the missing data mechanism concerning the outcome variable, Y_i . In particular, we assume that the pattern of missing data is conditionally independent of potential outcomes given the compliance status, the treatment status, and the observed covariates (Frangakis and Rubin, 1999),

$$R_i \perp \{Y_i(0), Y_i(1)\} \mid C_i, T_i, X_i \quad \text{for all } i, \quad (4)$$

where \perp denotes independence.³⁶ That is, by using C_i , T_i , and X_i , we predict the missing values of the outcome variable via the model specified in equation 3.³⁷

The key here is to include important predictors of the outcome variable in X_i so that the above assumption holds. Such variables also lead to an efficient estimation of causal effects because they help to accurately predict the missing values of compliance status and outcome variable. This yields the first of our four methodological recommendations made in Section 1. In our experiment, along with many background characteristics that are known to be strongly associated with the voting behavior of Japanese electorate, X_i includes voting intention variables that are measured two weeks prior to the election. Therefore, the assumption seems quite reasonable. In other experiments, researchers might also measure the past values of the outcome variable. It is also important to emphasize that our pre-election survey is designed to ensure that there is no missing

³⁵These points apply more generally to any analysis with missing data (King, Honaker, Joseph, and Scheve, 2001; Rubin, 1976).

³⁶This assumption is called the latent ignorability because the missing mechanism depends on the partially observed variable, C_i (Frangakis and Rubin, 1999).

³⁷In order to control for pre-treatment covariates, it is also possible to combine methods of matching and sub-classification with the model-based approach presented here (see Ho, Imai, King, and Stuart, 2004; Imai and van Dyk, 2004).

value among the pre-treatment variables. Although having a set of fully observed pre-treatment covariates is useful, in general researchers may encounter missing values for both the pre-treatment and response variables. The statistical framework and model we present in this section can be extended to such situations (see e.g., Barnard *et al.*, 2003).

The two probit models in equations 2 and 3 are combined to form the following complete-data likelihood function,

$$\prod_{i=1}^N \left[\Phi(X_i^\top \xi) \Phi(\alpha T_i + \beta(1 - T_i) + X_i^\top \gamma)^{Y_i} \{1 - \Phi(\alpha T_i + \beta(1 - T_i) + X_i^\top \gamma)\}^{1 - Y_i} \right]^{C_i} \times \left[\{1 - \Phi(X_i^\top \xi)\} \Phi(X_i^\top \gamma)^{Y_i} \{1 - \Phi(X_i^\top \gamma)\}^{1 - Y_i} \right]^{1 - C_i}. \quad (5)$$

We cannot directly evaluate this likelihood function for two reasons. First, the compliance status, C_i , is not observed for the voters in the control group. Second, the value of the outcome variable, Y_i , is missing for some observations. In situations such as this, it is natural to consider the method of data augmentation – a fundamental idea of Bayesian statistics – where missing data are “imputed” based on statistical models (Tanner and Wong, 1987; van Dyk and Meng, 2001). In our case, each iteration of our estimation procedure imputes the unobserved values of C_i and Y_i using the probit models with estimated coefficients and the pre-treatment variables of respondent i . Based on the imputed values, we then update the estimated coefficients. We repeat these steps until a satisfactory degree of convergence is achieved.

To conduct the Bayesian analysis, we assign two independent conjugate prior distributions on (α, β, γ) and ξ , both of which are multivariate normal distributions with mean zero and large variance. To estimate the model, we sample from the joint posterior distribution via a Gibbs sampling algorithm. We use the marginal data augmentation algorithm of Imai and van Dyk (2005) to exploit the latent variable structure and speed up the convergence. The details of the algorithm appear in Appendix B. The Gibbs sampling algorithm produces Monte Carlo samples from the joint posterior distribution of the unobserved compliance status, the missing values of the outcome variable, and the model parameters. Given these posterior draws, the estimates of CACE

and the ITT effect can be easily calculated, as can estimates of the uncertainty in those estimates. Our convergence diagnostics are based on multiple independent Markov chains initiated at over-dispersed starting values (Gelman and Rubin, 1992). The sampling algorithm is implemented by our own C code, which we make publicly available with an easy-to-use R interface.

5 Estimated Causal Effects of Information on Voter Turnout

In this section, we present the results of our statistical analysis. We use independent and diffuse prior distributions; i.e., normal distribution with a mean of zero and a variance of 100 for each of the coefficients of both the compliance and outcome models. Our inference is based on Monte Carlo samples from three independent Markov chains, each of which has the length of 20,000 and is initiated at different sets of over-dispersed starting values.³⁸ We find that all the parameters have the values of the Gelman-Rubin convergence statistic that are less than 1.01, which suggests that a satisfactory degree of convergence has been achieved. We retain the last 10,000 draws from each chain and base our inference on a combined total of 30,000 posterior draws.

5.1 One-party versus Two-party Treatment Effects

First, Table 2 compares the estimated causal effect of one-party treatment with that of two-party treatment by presenting the posterior summaries of quantities of interest; the ITT effect, CACE, and proportion of compliers in the sample. Also, the estimated turnout rates for the control group are shown separately for compliers, noncompliers, and the entire sample. They serve as the baseline turnout of no exposure to the designated party websites.³⁹ Note that compliers

³⁸For the first chain, we use zero as the starting value for all the model parameters. For the second chain, we use 5 for all the coefficients of the compliance model and -5 for all the coefficients of the outcome model. For the third chain, we use -5 for the compliance model and 5 for the outcome model.

³⁹These turnout rates for the control group need to be estimated because we do not observe the outcome variable for some voters and the compliance status is not known for every voter in the control group. As a result, the model

	Summary of posterior distributions			
	Mean	s.d.	2.5%	97.5%
One-party treatment				
Intention-to-treat (ITT) effect	0.010	0.022	-0.033	0.052
Complier average causal effect (CACE)	0.013	0.029	-0.044	0.070
Fraction of compliers	0.751	0.007	0.736	0.764
Turnout for the control group				
Compliers	0.705	0.023	0.661	0.750
Noncompliers	0.678	0.057	0.563	0.785
All	0.698	0.009	0.680	0.715
Two-party treatment				
Intention-to-treat (ITT) effect	0.033	0.023	-0.012	0.077
Complier average causal effect (CACE)	0.047	0.032	-0.017	0.110
Fraction of compliers	0.703	0.012	0.680	0.726
Turnout for the control group				
Compliers	0.683	0.027	0.631	0.737
Noncompliers	0.730	0.054	0.620	0.829
All	0.697	0.009	0.678	0.715

Table 2: Estimated Causal Effects of Policy Information on Voter Turnout for One-party and Two-party Treatments. The figures represent the numerical summaries of posterior distributions for each quantity of interest separately for the one-party treatment and two-party treatment conditions: standard deviation, and 95 percent credible interval. The estimated turnout rates for the control group are shown separately for compliers, noncompliers, and the entire sample, as the baseline turnout of no exposure to the designated party websites. Each model produces slightly different estimates of turnout rates for the control group. Moreover, compliers are defined separately for each treatment.

are defined separately for each treatment; i.e., compliers for the one-party treatment are not necessarily the same as compliers for the two-party treatment. The estimated baseline turnout rate is about 70 percent, which is almost 15 percentage points higher than the official turnout rate of the 2004 Upper House election. This gap arises because our sample is not representative of the Japanese electorate.⁴⁰ The estimated model parameters for the two-party treatment effect appear

for one-party treatment produces the estimates of turnout rates for the control group that are slightly different from the model for two-party treatment.

⁴⁰ It is also possible that the self-reported turnout is biased (e.g., Burden, 2000; Campbell *et al.*, 1960; Silver *et al.*, 1986). The magnitude of self-reporting bias is minimal unless the degree of misreporting is affected by the actual treatment or the treatment assignment.

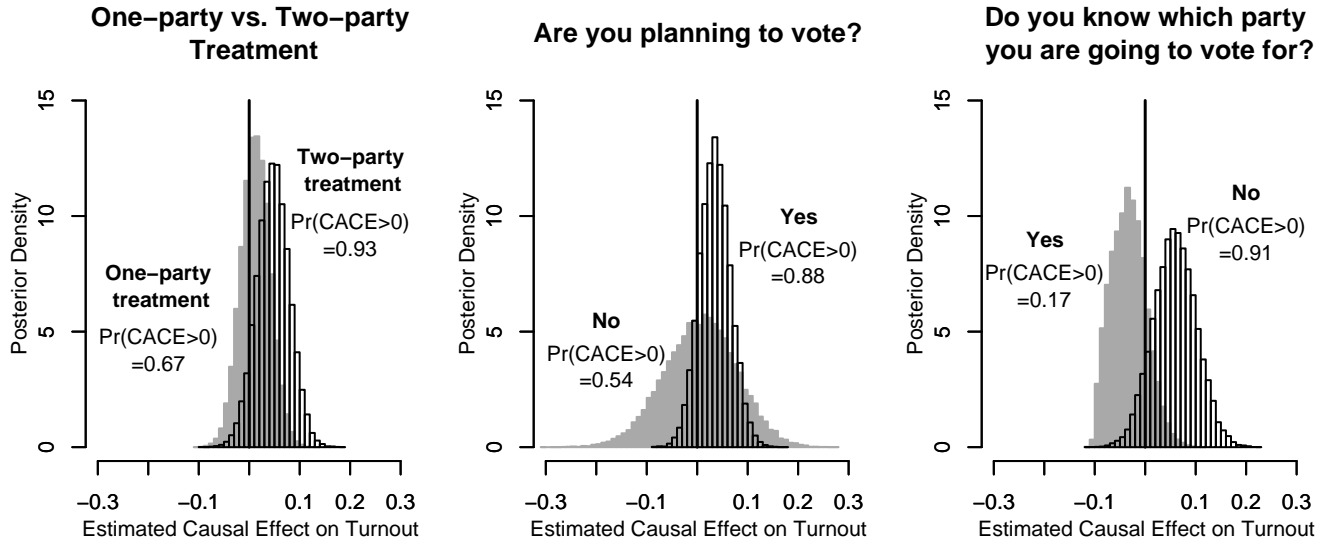


Figure 3: Histograms of Posterior Simulation Draws for the Estimated Complier Average Causal Effects (CACE) on Turnout. The left panel compares the one-party treatment effect (shaded histogram) with the two-party treatment effect (unshaded histogram). The middle panel compares CACE using two subgroups – those voters who were planning to vote (unshaded) and those who were not (shaded). The right panel compares the causal effects for another set of two subgroups – those who knew which party they were going to vote for (shaded) and those who did not (unshaded). The vertical lines represent the zero causal effect. The posterior probability that CACE exceeds zero is also shown for each subgroup.

in Appendix C.

As expected from the hypothesis, the posterior means of CACE and the ITT effect are all positive. The estimated causal effects of one-party treatment are small and there is relatively large uncertainty. The estimated probability of voting increases (from 70.5 percent) by only 1.3 percentage points on average among those voters who *actually visited* one party website (CACE), and the estimated increase is on average one percentage point (from 69.8 percent) among those who were *asked to visit* one website (the ITT effect).⁴¹ Since the estimated proportion of compliers is quite high (75.1 percent in the one-party treatment and 70.3 percent in the two-party treatment), the estimated ITT effects and CACE are somewhat similar in our study. Large posterior standard deviations mean that one cannot statistically distinguish these small positive estimates from zero.

⁴¹The estimated CACE for visiting the LDP website is somewhat larger than that of viewing the DPJ website though the uncertainty estimates become even larger.

In contrast, the estimated causal effects of the two-party treatment are larger. The estimated turnout probability increases (from 69.7 percent) by 3.3 percentage points on average if a voter is asked to visit the websites of both the LDP and the DPJ. The estimated causal effect among compliers is even larger, showing that those who actually visited the two party websites were on average 4.7 percentage points more likely to vote than those who did not. Although the 95 percent Bayesian credible intervals include zero for both ITT effect and CACE, approximately 93 percent of posterior draws take positive values. This suggests that exposing voters to the websites of both parties has a positive effect on their turnout. The left panel of Figure 3 presents the histograms of posterior simulation draws for one-party and two-party treatments, and graphically illustrates the difference between the two.

We note that the comparison of the two CACE estimates requires some caution because the compliers of different treatments might differ in their characteristics.⁴² Nevertheless, our findings offer support for the hypothesis that additional information increases voter turnout, especially when voters are exposed to the policy information of *both* ruling and opposition parties (as opposed to just one party). One possible explanation is that voters can better understand policy differences of the two parties when they compare their actual policy proposals, and this makes them more likely to vote. Future research should theoretically examine and empirically investigate different causal mechanisms that explain this difference between one-party and two-party treatment effects.

One may argue that the estimated positive effects merely reflect the participation in the survey and are not attributable to the exposure to policy information. To investigate this possibility, we conduct a sensitivity analysis by redefining those voters who have logged onto the pre-election survey as the treated units. This definition includes the additional 133 voters who started the survey but did not visit the designated website. As expected, we find that the resulting estimated causal effects are smaller by approximately 25 percent. For the two-party treatment effect, for

⁴²This problem is severe when the number of noncompliers is large. The comparison of two estimated ITT effects, on the other hand, is straightforward.

example, we estimate 2.7 percentage points for the ITT effect (posterior standard deviation 2.2) and 3.5 percentage points for CACE (posterior standard deviation 2.8) on average. This suggests that visiting the designated party websites, rather than participating in the survey, increases turnout. Moreover, regardless of definitions of actual treatment status, the estimated effect of one-party treatment is smaller than that of two-party treatment. This indicates that different quantities of policy information yield varying degrees of increase in voter turnout. We emphasize that our ability to conduct this kind of detailed analysis depends on the precise measurement of actual treatment status we obtained in our pre-election survey. This point is explicitly made in our third methodological recommendation in Section 1.

5.2 Voting Intention and Causal Effects of Policy Information

Next, we examine whether the size of causal effects differs across certain types of voters. Using some of the pre-treatment variables measured in the screening survey, we estimate causal effects separately for different subgroups that are defined by voting intention. Table 3 presents the results of our analysis for four subgroups of interest. Because of the limited sample sizes, we estimate the effect of visiting at least one party’s website by pooling the one-party and two-party treatment groups. We compare those voters who said they were planning to vote with those who said they were undecided or not planning to vote, using one of the pre-treatment variables used to define randomized blocks.⁴³ The results indicate that visiting at least one designated website increases turnout by 3.5 percentage points (from 86.0 percent) on average among those who said they were planning to vote. However, we find very little effect for those who said they were undecided or not planning to vote. While the small sample size makes the finding somewhat inconclusive, our treatment had little effect on those who did not have a strong intention to vote in the first place. We consider the effect size quite large for those voters who were planning to vote given that the

⁴³We pool undecided voters and those who were not planning to vote because of limited sample sizes for these two groups.

	Summary of posterior distributions			
	Mean	s.d.	2.5%	97.5%
Planning to vote				
Intention-to-treat (ITT) effect	0.026	0.023	-0.018	0.072
Complier average causal effect (CACE)	0.035	0.031	-0.024	0.096
Fraction of compliers	0.750	0.006	0.738	0.761
Turnout for the control group				
Compliers	0.860	0.022	0.819	0.904
Noncompliers	0.850	0.052	0.741	0.941
All	0.858	0.009	0.838	0.874
Undecided/Not planning to vote				
Intention-to-treat (ITT) effect	0.004	0.049	-0.095	0.099
Complier average causal effect (CACE)	0.006	0.072	-0.138	0.143
Fraction of compliers	0.694	0.009	0.674	0.712
Turnout for the control group				
Compliers	0.315	0.056	0.205	0.424
Noncompliers	0.369	0.090	0.200	0.550
All	0.333	0.022	0.293	0.374
Knew which party they were going to vote for				
Intention-to-treat (ITT) effect	-0.025	0.026	-0.070	0.030
Complier average causal effect (CACE)	-0.033	0.034	-0.092	0.039
Fraction of compliers	0.768	0.008	0.753	0.783
Turnout for the control group				
Compliers	0.931	0.025	0.883	0.974
Noncompliers	0.885	0.064	0.757	1.000
All	0.920	0.010	0.897	0.938
Didn't know which party they were going to vote for				
Intention-to-treat (ITT) effect	0.042	0.030	-0.017	0.101
Complier average causal effect (CACE)	0.059	0.043	-0.024	0.142
Fraction of compliers	0.713	0.006	0.700	0.725
Turnout for the control group				
Compliers	0.555	0.033	0.491	0.621
Noncompliers	0.605	0.065	0.475	0.728
All	0.570	0.013	0.543	0.594

Table 3: Voting Intention and Estimated Causal Effects of Policy Information on Voter Turnout. The figures represent the numerical summaries of posterior distributions for each quantity of interest separately for the one-party treatment and two-party treatment conditions: standard deviation, and 95 percent credible interval. The estimated turnout rates for the control group within each subsample are shown separately for compliers, noncompliers, and the entire subsample, as the baseline turnout of no exposure to the designated party websites.

baseline turnout without exposure to the party websites is estimated to be greater than 85 percent. The middle panel of Figure 3 shows the histograms of posterior distributions of CACE for the two subgroups. Approximately 88 percent of posterior draws take positive values for the subgroup of those who said they were planning to vote.

We also compare the group of voters who knew which party they were going to vote for and those who did not know two weeks prior to the election.⁴⁴ The results indicate that visiting at least one designated party website has a slightly negative effect on the former group. In fact, 83 percent of posterior draws for CACE take negative values. In contrast, the policy information raises turnout by 5.9 percentage points on average among those who did not know which party they were going to vote for in the election. Again, although a relatively large posterior standard deviation prevents us from drawing a definitive conclusion, the right panel of Figure 3 shows that 91 percent of posterior draws are positive for this subgroup. While our experimental sample is too small to conduct a direct test and draw a definitive conclusion, the independent analyses of these subgroups suggest that policy information may have a large positive impact on the turnout of those who are planning to vote, and yet are undecided about which party to vote for in the election.

6 Concluding Remarks

In this paper, we show how to effectively design and analyze randomized experiments using our Japanese election experiment as an example. First, in order to ensure efficient randomization, we recommend the use of randomization schemes such as randomized block and matched-pair designs. Second, the Bayesian statistical method used in this paper can overcome the problems of noncompliance and nonresponse, which are frequently encountered in randomized experiments.

⁴⁴For this analysis, the convergence of three chains is slower. Therefore, we run the chains longer (50,000 draws instead of 20,000) and keep the last 10,000 draws from each chain.

Third, by measuring a number of important pre-treatment covariates and minimizing the number of noncompliers and nonresponses, researchers can make modeling assumptions more plausible and therefore enhance the validity of causal inference. Fourth, precise measurement of treatment status is essential for accurate interpretation of estimated causal effects. We believe that our methodological recommendations are widely applicable to other experimental studies. We also hope that our Japanese election experiment serves as a methodological template for future causal inquiry with Internet survey-based randomized experiments.

The development of statistical methods for causal inference with experimental data is an active area of research. Here, we identify some of the important remaining methodological challenges for future research. First, while this paper dealt with binary treatment variables, the methods need to extend to situations where the treatment variable is ordered categorical, multinomial or even continuous (e.g., Imai and van Dyk, 2004). Second, the model used in this paper essentially relies on the constant treatment effect assumption. However, causal effects are likely to vary across individuals. More sophisticated statistical models must be developed in order to address the heterogeneity of treatment effects (e.g., Angrist, 2004). Finally, it is important to consider methods that use the estimated causal effects from a specific and nonrandom sample to predict the outcomes in different samples (e.g., Hotz, Imbens, and Mortimer, 2005). Such methods are of great interest to practitioners deciding future policies based on program evaluations. They are also useful for researchers who wish to draw conclusions about a relevant population using experiments conducted with a somewhat unrepresentative sample.

Appendices

A Balance of Observed Covariates After Randomization

The figures in Table 4 represent t -statistics testing mean differences for each pretreatment variable across six randomized blocks. Each of the four columns shows the overall balance of the variables separately for each treatment group compared with the control group. No t -statistic indicates statistically significant mean differences at the 0.05 level, implying that the pretreatment variables are well balanced for all treatment assignments.

B Computational Details

This appendix gives details of the Gibbs sampling algorithm that we use to fit the model proposed in Section 4.3. We begin the algorithm with the starting value for parameters $(\alpha^{(0)}, \beta^{(0)}, \gamma^{(0)}, \xi^{(0)})$ and missing data $(C^{(0)}, Y^{(0)})$. The compliance status for the units in the treatment assignment group is known; i.e., $C_i^{(0)} = C_i = T_i$ for units with $Z_i = 1$. Similarly, we set $Y_i^{(0)} = Y_i$ for units with $R_i = 1$. We then proceed via the following three steps at iteration t ,

Step 1: Sample the binary compliance status $C_i^{(t)}$ independently for each i with $Z_i = 0$ from the Bernoulli distribution with probability,

$$\frac{\lambda_i [Y_i \omega_i + (1 - Y_i)(1 - \omega_i)]}{\lambda_i [Y_i \omega_i + (1 - Y_i)(1 - \omega_i)] + (1 - \lambda_i) [Y_i \pi_i + (1 - Y_i)(1 - \pi_i)]},$$

where $\lambda_i = \Phi(X_i^\top \xi^{(t-1)})$, $\pi_i = \Phi(X_i^\top \gamma^{(t-1)})$, and $\omega_i = \Phi(\beta^{(t-1)} + X_i^\top \gamma^{(t-1)})$. For units with $Z_i = 1$, set $C_i^{(t)} = C_i^{(t-1)}$.

Step 2: Impute the missing values of the outcome variable for each i with $R_i = 0$ by sampling from the Bernoulli distribution with probability $\Phi(\alpha^{(t-1)} C_i^{(t)} T_i + \beta^{(t-1)} C_i^{(t)} (1 - T_i) + X_i^\top \gamma^{(t-1)})$. For units with $R_i = 1$, set $Y_i^{(t)} = Y_i^{(t-1)}$.

Step 3: Given the updated compliance status $C_i^{(t)}$, perform the Bayesian probit regression for the compliance model using the marginal data augmentation scheme in Section 3.3 of Imai and van Dyk (2005). This gives the new draws of the model parameter, $\xi^{(t)}$.

Step 4: Given the updated outcome variable $Y_i^{(t)}$, perform the Bayesian probit regression for the outcome model again using the marginal data augmentation scheme in Section 3.3 of Imai and van Dyk (2005). This gives the new draws of the model parameters, $(\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)})$.

C Estimated Model Parameters

Table 5 shows the posterior summaries for each of the estimated parameters of compliance and outcome models in our Bayesian model with the two-party treatment. The covariates used in the estimation include each respondent's age and its square, whether her highest education completed is college or above, whether she knew which party they were going to vote for in the 2004 Upper House election (four point scale), and whether she thought the LDP, New Kōmeitō, JCP (Japan Communist Party), or SDP (Social Democratic Party) was her preferred party (four point scale). These variables are taken from the screening survey. We also added the aggregate voter turnout rate in the prefecture of the respondent's residence in the previous Upper House election in 2001.

	One party treatment		Two party treatment	
	LDP	DPJ	LDP/DPJ	DPJ/LDP
Age	-0.27	0.09	-0.97	-1.11
Age ²	-0.19	0.09	-0.89	-1.04
Gender (Male=1, Female=2)	0.06	0.06	0.09	0.09
Highest education				
Junior highschool	-0.27	-0.68	-0.13	0.35
Highschool	0.39	-0.47	0.46	0.57
Vocational school	-1.21	-0.12	0.44	-0.39
Two-year college	-0.44	0.47	-0.09	0.85
College	1.27	0.43	-0.20	-1.00
Graduate school	-1.09	-0.43	-0.85	0.09
Four point scale variables about voter preferences				
Preferred party or candidate	0.14	-0.27	-0.32	-0.28
LDP is the preferred party	-0.77	0.26	0.24	0.39
DPJ is the preferred party	-0.24	0.90	-0.30	-0.18
Kōmeitō is the preferred party	-0.21	0.64	0.77	0.50
JCP is the preferred party	-0.27	1.00	-0.30	0.83
SDP is the preferred party	0.37	1.23	-0.78	-0.46
Planning to vote	-0.08	-0.08	0.02	0.02
Proportional representation: Planning to vote for				
LDP	0.24	0.49	-1.31	-0.79
DPJ	-0.08	-1.16	-0.14	-0.70
Kōmeitō	0.35	0.32	0.61	0.87
JCP	0.93	1.57	0.81	0.46
SDP	-0.26	-0.95	0.36	0.66
Other party	-1.23	0.19	0.75	0.83
Confidence (four point scale)	-0.60	0.04	0.19	0.17
Electoral district: Planning to vote for				
LDP candidate	0.99	0.89	-1.35	-0.06
DPJ candidate	-0.22	-0.46	-0.08	-0.46
Kōmeitō candidate	-0.62	-0.29	0.15	-0.51
JCP candidate	0.11	0.66	0.64	-0.85
SDP candidate	-1.49	-1.80	-0.34	-1.08
Independent candidate	-0.70	-0.31	0.38	1.29
Other candidate	1.66	0.80	1.81	1.35
Confidence (four point scale)	0.68	0.08	0.20	0.44
Not planning to vote	0.07	0.07	0.17	0.17
Undecided	0.04	0.04	-0.17	-0.17

Table 4: Overall Balance of Pretreatment Variables Across Randomized Blocks. The figures represent t -statistics for all pre-treatment covariates that compare each treatment group with the control group.

	Summary of posterior distributions			
	Mean	s.d.	Mean	s.d.
	Compliance model		Outcome model	
Compliers receiving the treatment			-0.203	0.188
Compliers not receiving the treatment			-0.416	0.270
Intercept for Block I	2.132	1.313	0.323	1.363
Intercept for Block II	2.013	1.315	0.005	1.357
Intercept for Block III	2.224	1.329	-1.670	1.363
Intercept for Block IV	1.884	1.310	-1.676	1.356
Intercept for Block V	1.585	1.318	-1.009	1.346
Intercept for Block VI	2.107	1.323	-0.907	1.363
Age/10	-0.673	0.439	-0.018	0.479
Age ² /100	0.087	0.058	0.035	0.065
Highest education completed: college or above	0.022	0.116	0.029	0.121
Knew which party they were going to vote for	0.144	0.131	0.548	0.146
LDP is the preferred party	-0.029	0.133	-0.193	0.138
DPJ is the preferred party	0.016	0.142	0.163	0.148
New Kōmeitō is the preferred party	0.107	0.162	-0.097	0.165
JCP is the preferred party	-0.295	0.165	-0.211	0.168
SDP is the preferred party	0.173	0.165	0.199	0.174
Aggregate turnout in 2001	-0.006	0.017	0.010	0.018

Table 5: Posterior Summaries for Each of the Estimated Parameters in the Two-Party Effect Model. The columns show the mean and standard deviation of the posterior distributions.

References

- Alvarez, R. M. (1998). *Information and Elections: Revised to Include the 1996 Presidential Election*. University of Michigan Press, Ann Arbor, MI.
- Angrist, J. D. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal* **114**, C52–C83.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association* **91**, 444–455.
- Ansolabehere, S. and Iyengar, S. (1995). *Going Negative: How Political Advertisements Shrink and Polarize the Electorate*. Free Press, New York, NY.
- Barnard, J., Frangakis, C. E., Hill, J. L., and Rubin, D. B. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York (with discussion). *Journal of the American Statistical Association* **98**, 462, 299–311.
- Bartels, L. M. (1996). Uninformed votes: Information effects in presidential elections. *American Journal of Political Science* **40**, 1, 194–230.
- Burden, B. C. (2000). Voter turnout and the national election studies. *Political Analysis* **8**, 4, 389–398.
- Campbell, A., Converse, P. E., Miller, W. E., and Stokes, D. E. (1960). *The American Voter*. John Wiley & Sons, New York, NY.
- Clinton, J. D. and Lapinski, J. S. (2004). “Targeted” advertising and voter turnout: An experimental study of the 2000 presidential election. *Journal of Politics* **66**, 1, 69–96.
- Cox, D. R. (1958). *Planning of Experiments*. John Wiley & Sons, New York.

- Cox, D. R. and Reid, N. (2000). *The Theory of the Design of Experiments*. Chapman & Hall, New York.
- Cox, G. W., Rosenbluth, F. M., and Thies, M. F. (1998). Mobilization, social networks, and turnout: evidence from Japan. *World Politics* **50**, 3, 447–474.
- Downs, A. (1957). *An Economic Theory of Democracy*. HarperCollins, New York, NY.
- Dulio, D. A., Goff, D. L., and Thurber, J. A. (1999). Untangled web: Internet use during the 1998 election. *PS: Political Science and Politics* **32**, 1, 53–59.
- Feddersen, T. and Pesendorfer, W. (1999). Abstention in elections with asymmetric information and diverse preferences. *American Political Science Review* **93**, 2, 381–398.
- Ferejohn, J. A. and Kuklinski, J. H., eds. (1990). *Information and Democratic Processes*. University of Illinois Press, Urbana and Chicago, IL.
- Frangakis, C. E. and Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86**, 365–379.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulations using multiple sequences (with discussion). *Statistical Science* **7**, 457–472.
- Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review* **94**, 653–663.
- Ghirardato, P. and Katz, J. N. (2002). Indecision theory: Quality of information and voting behavior. Social Science Working Paper 1106R, California Institute of Technology.
- Gibson, R., Nixon, P., and Ward, S., eds. (2003a). *Political Parties and the Internet: Net Gain?* Routledge, London.

- Gibson, R. K., Margolis, M., Resnick, D., and Ward, S. J. (2003b). Election campaigning on the WWW in the USA and UK: A comparative analysis. *Party Politics* **9**, 1, 47–75.
- Greevy, R., Lu, B., Silber, J. H., and Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics* **5**, 263–275.
- Grofman, B., ed. (1993). *Information, Participation, and Choice: An Economic Theory of Democracy in Perspective*. University of Michigan Press, Ann Arbor, MI.
- Hill, J., Rubin, D. B., and Thomas, N. (1999). *Research Designs: Inspired by the Work of Donald Campbell* (eds. L. Bickman), chap. The Design of the New York School Choice Scholarship Program Evaluation, 155–180. Sage, Thousand Oaks, CA.
- Hirano, K., Imbens, G. W., Rubin, D. B., and Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1**, 69–88.
- Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2004). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Working Paper* available at <http://www.princeton.edu/~kimai/research/preprocess.html>.
- Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association* **81**, 945–960.
- Hotz, V. J., Imbens, G. W., and Mortimer, J. H. (2005). Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics* **125**, 1–2, 241–270.
- Huckfeldt, R. R. and Sprague, J. (1995). *Citizens, Politics, and Social Communication: Information and Influence in An Election Campaign*. Cambridge University Press, New York, NY.
- Imai, K. (2005). Do get-out-the-vote calls reduce turnout?: The importance of statistical methods for field experiments. *American Political Science Review* **99**, 2, 283–300.

- Imai, K. and van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association, Theory and Methods* **99**, 467, 854–866.
- Imai, K. and van Dyk, D. A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics* **124**, 2, 311–334.
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* **86**, 1, 4–29.
- Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25**, 305–327.
- Iyengar, S. and Jackman, S. (2003). Can information technology energize voters? experimental evidence from the 2000 and 2002 campaigns. Presented at the Annual Meeting of the American Political Science Association, Philadelphia, August, 2003.
- Iyengar, S. and Kinder, D. R. (1987). *News That Matters: Television and American Opinion*. University of Chicago Press, Chicago, IL.
- Johnson, T. J. and Kaye, B. K. (2003). A boost or bust for democracy? how the web influenced political attitudes and behaviors in the 1996 and 2000 presidential elections. *Harvard International Journal of Press/Politics* **8**, 3, 9–34.
- Kinder, D. R. and Palfrey, T. R., eds. (1993). *Experimental Foundations of Political Science*. University of Michigan Press, Ann Arbor, MI.
- King, G., Honaker, J., Joseph, A., and Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review* **95**, 1, 49–69.

- Lassen, D. D. (2005). The effect of information on voter turnout: Evidence from a natural experiment. *American Journal of Political Science* **49**, 1, 103–118.
- Lupia, A. (1994). Shortcuts versus encyclopedias: Information and voting behavior in California insurance reform elections. *American Political Science Review* **88**, 1, 63–76.
- Lupia, A. and Philpot, T. S. (Nd). Views from inside the net: How websites affect young adults’s political interest. *Journal of Politics* Forthcoming.
- Matusaka, J. G. (1995). Explaining voter turnout patterns: An information theory. *Public Choice* **84**, 1–2, 91–117.
- McDermott, R. (2002). Experimental methods in political science. *Annual Review of Political Science* **5**, 31–61.
- Mutz, D. C. (2002). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review* **96**, 1, 111–126.
- Palfrey, T. R. and Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review* **79**, 1, 62–78.
- Pharr, S. J. (1981). *Political Women in Japan: The Search For a Place in Political Life*. University of California Press, Berkeley, CA.
- Rosenstone, S. J. and Hansen, J. M. (1993). *Mobilization, Participation, and Democracy in America*. Macmillan, New York, NY.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* **66**, 688–701.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

- Rubin, D. B. (1990). Comments on “On the application of probability theory to agricultural experiments. Essay on principles. Section 9” by J. Splawa-Neyman translated from the Polish and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science* **5**, 472–480.
- Sekhon, J. (2004). The varying role of voter information across democratic societies. *Working Paper* Department of Government, Harvard University.
- Silver, B. D., Anderson, B. A., and Abramson, P. R. (1986). Who overreports voting? *American Political Science Review* **80**, 2, 613–624.
- Smith, A. F. M. (1973). Bayes estimates in one-way and two-way models. *Biometrika* **60**, 319–329.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528–550.
- van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation (with discussions). *The Journal of Computational and Graphical Statistics* **10**, 1–111.
- Wantchekon, L. (2003). Clientelism and voting behavior: Evidence from a field experiment in Benin. *World Politics* **55**, 399–422.
- Weber, L. M., Loumakis, A., and Bergman, J. (2003). Who participates and why?: An analysis of citizens on the Internet and the mass public. *Social Science Computer Review* **21**, 1, 26–42.