

**Experimenting in Democracy Promotion:
International Observers and the 2004 Presidential Elections in Indonesia***

Susan D. Hyde
Assistant Professor of Political Science and International Affairs
Yale University
(203)432-5672 (office)
(203) 432-3296 (fax)
susan.hyde@yale.edu

* I wish to thank Jeffrey Isaac and two anonymous reviews for their extremely helpful comments, as well as Eric Bjornlund, David Carroll, Avery Davis-Roberts, Clark Gibson, Vladimir Pran, and David Pottie for making this study possible and for providing useful feedback on multiple occasions. Thanks are also due to Thad Dunning, Don Green, David Lake, Mat McCubbins, and Irfan Nooruddin for numerous comments on the project, and to Gary Cox, J. Morgan Kousser, and participants at seminars at The Ohio State, Princeton, Columbia, and New York University. Sarah Bush provided valuable research assistance, and Jon Immel provided useful technical assistance. Any errors or omissions are my own.

**Experimenting in Democracy Promotion:
International Observers and the 2004 Presidential Elections in Indonesia**

Abstract

Randomized field experiments have gained attention within the field of democracy promotion and within the social sciences as an influential tool for causal inference and a potentially powerful method of impact evaluation. With an eye towards facilitating field experimentation in democracy promotion, this article presents the first field-experimental study of international election monitoring, an application of field experimental methods that should be of interest to both practitioners and academics. I use the random assignment of international observers during election-day observation of the 2004 presidential elections in Indonesia to illustrate how experiments can be used to evaluate the effects monitoring on election-day behavior. The study demonstrates that international observers can influence even democratic elections, and underscores that experiments have a unique ability to reveal unanticipated effects of democracy promotion programs.

International election observers are now present at more than four out of every five elections in the developing world. International monitoring of domestic elections was once regarded as an unacceptable violation of state sovereignty.¹ Since the late 1990s, the refusal by a government to invite reputable foreign observers has become a conspicuous signal that democratic elections are unlikely to take place.² Despite the widespread presence of election observers throughout the world, their effects on elections are not fully understood. Do international observers directly influence the quality of elections? Or does the presence of foreign observers have little or no effect on voters, parties, and election officials? In this article I present field experimental methods and the random assignment of election observers as one way to answer this question. More generally, I discuss field experiments as a promising method for evaluating the effects of democracy assistance programs. Applied to the 2004 presidential elections in Indonesia, the random assignment of international election observers reveals that even though the election was widely regarded as democratic, the presence of observers had a measurable effect on votes cast for the incumbent candidate, indicating that such democracy assistance can influence election quality even in the absence of blatant election day fraud.

Additionally, because this application of field experimental methods was the first of its kind in election monitoring, I present this study with an eye toward facilitating future research and debate about the implementation of field experiments in what are often challenging conditions. These conditions frequently involve short time horizons, limited access to relevant outcome measures, and potentially large but uncertain payoffs.

Field Experiments and the Effects of Democracy Promotion

Democracy promotion's history is long and diverse, significantly pre-dating the introduction of election monitoring in sovereign states, and going back at least to the early

twentieth century.³ Following both the end of World War II and the end of the Cold War, increasing numbers of democratic states and international organizations incorporated democracy promotion explicitly into their policies toward other countries.⁴ Various forms of foreign assistance from states, NGOs, intergovernmental organizations, and international financial institutions became conditioned on democracy and democratization. These international actors became more likely to reduce support or sanction countries experiencing military coups, failing to hold democratic elections, or engaging in massive violations of political rights. Direct democracy assistance became a growth industry, and developed democracies, international NGOs, or pro-democracy intergovernmental organizations like the European Union, the Organization of American States, or the Organization for Security and Cooperation in Europe devoted significant resources to direct democracy assistance.

Despite the billions of dollars spent on democracy assistance since the early 1990s, and the prominent role that democracy promotion plays in the foreign policies of many influential governments, ongoing uncertainty about the conditions under which democracy assistance “works” continues to fuel debate among scholars and policy makers. On one hand, critics argue that democracy assistance is imperialist and interventionist, imposed on unwilling populations who would be better off without such foreign meddling, or as a disingenuous cover for other more controversial foreign policy objectives.⁵ On the other hand, proponents argue that democracy assistance and democracy promotion more generally have provided crucial support for democratizing forces during critical junctures in the political development of a number of well known cases.⁶ Additionally, individual governments tout their foreign policy successes in democracy promotion in terms of dollars spent on democracy assistance or highlight the correlation between their work and successful democratic transitions.⁷ Because they have a

vested interest in continuing their own programs, practitioners of democracy promotion are frequently less credible as evaluators of democracy assistance programs.⁸

Missing from much of the debate over democracy assistance is a more objective method by which scholars and practitioners can evaluate the conditions under which various forms of democracy assistance are effective. The question of whether and how international actors influence the domestic politics of sovereign states is of interest for academic and policy-related reasons. Correlations between total dollars spent on democracy assistance and changes in aggregate measures of democracy or political rights are suggestive but somewhat controversial in that they are open to numerous interpretations. Because democracy promoters and aid donors may target their assistance toward regimes where improvements are most likely (or least likely), it is difficult to separate the consequences of democracy promotion from what would have happened in the country in the absence of the democracy assistance programs. Therefore, although it is possible to show that average levels of democracy improve when more money is spent on democracy assistance, it is extremely difficult to demonstrate that the change in democracy scores was caused by democracy assistance rather than some other omitted variable. It is even more difficult to use such methods to pinpoint which democracy assistance programs work as they were intended, which programs do not have their intended effects, and whether any programs have unanticipated effects.

Although many scholars are circumspect about the conclusions that can be drawn from such research, individuals on both sides of the debate on democracy promotion have been guilty of asserting a causal relationship between democracy assistance and either positive or negative outcomes. For example, as Finkel et al. argue in their large-N cross national study of democracy aid, “there are consistent positive impacts of direct USAID democracy assistance on overall

levels of democracy in recipient countries, as measured by the Freedom House and Polity IV indices over time.”⁹ Other scholars have found positive, negative and insignificant relationships between aggregate levels of democracy and money spent on democracy assistance, and the relationship between democracy and foreign aid.¹⁰ Some detailed case studies make convincing arguments that democracy promoters contributed to negative outcomes, such as international involvement in the 1992 elections in Angola or the 1993 elections in Cambodia.¹¹ Other scholars argue that democracy promoters were instrumental in bringing about democratic transitions or in preventing democratic backsliding, as in Peru in 2000 or more generally in post-Soviet Latin America and Central Europe.¹² Although it is clear that international efforts to promote democracy were followed either by disastrous outcomes (as in Angola and Cambodia) or desirable progress toward democracy (as in Peru after 2000 or Georgia in 2004), it is difficult to compare the actual outcome to the counterfactual outcome, or what would have happened if international actors had not attempted to promote democracy or pressure governments to hold democratic elections.

Field experiments offer a partial solution to several related problems, as a number of powerful players in the democracy promotion industry have recognized in recent years. The USAID commissioned National Academy of Sciences report, *Improving Democracy Assistance*, covers a wide range of methods for evaluating the effectiveness of democracy assistance programs, but recommends “randomized impact evaluation” as the “ideal research design.”¹³ Partly as a result, many recipients of USAID funding have begun to adopt such methods to evaluate their programs. Note that the NAS report does not recommend the exclusive use of field experiments, but highlights them as an underutilized and potentially powerful tool to help democracy promoters complete rigorous evaluations of pilot programs, learn about the effects of

existing programs, and ultimately refine the tools of democracy assistance over time. The Millennium Challenge Corporation, also funded by the US government, promotes democracy in the poorest countries as part of its mission, and calls for the “use of random assignment to treatment and control groups” as the method of impact evaluation likely to produce the most “rigorous results.”¹⁴

This article is therefore potentially important in part because of the new attention to “rigorous impact evaluation” spreading throughout the democracy promotion industry, and because it represents an example of the potential for mutually beneficial cooperation between academics and practitioners in studying the effects of democracy promotion.

International Election Observation

International election monitoring is one of the most well-known and potentially consequential forms of democracy promotion.¹⁵ Aid spent directly on democracy assistance activities (not including indirect democracy promotion like aid conditionality) represents hundreds of millions annually in the US alone, and over the course of the 1990s, election monitoring became a central part of the growing democracy promotion industry.¹⁶

Democratization—usually considered a purely domestic political process—since the end of the Cold War has become permeated with international actors.¹⁷ This development represents an overt attempt by international actors to influence the course of domestic politics, yet, like many such relationships, it remains understudied.¹⁸

States, IGOs, NGOs, and scholars who support election observation argue that it increases voter and political party confidence in the electoral process, deters fraud when it exists, and generates a third-party evaluation of election quality for international and domestic

audiences, thus making negative consequences for a leader who holds fraudulent elections more likely.¹⁹ As Kofi Annan stated while Secretary General of the United Nations:

The presence of international election observers, fielded always at the invitation of sovereign states, can make a big difference in ensuring that elections genuinely move the democratic process forward. Their mere presence can dissuade misconduct, ensure transparency, and inspire confidence in the process.²⁰

Skepticism of this view of observers is prevalent. The most critical argue that international election observers are simply glorified tourists,²¹ or that they are biased representatives of governments, out only to promote their country's narrow economic interests.²² Others argue that observers fail on other grounds, for example, that observers do not succeed in their mission when they actually observe electoral fraud because documenting fraud proves that they have failed to prevent it.²³ Even supporters of election observation suggest that observers should be more professionalized and receive better training, increase consistency in their evaluations across countries, improve coordination with local actors, and generally increase the accuracy of their evaluations of election quality.²⁴

An assortment of regional or case studies examine the role of international observers in the democratization process, and show mixed support for the claims made by proponents of election observation. For example, Bjornlund, Bratton and Gibson suggest that the presence of observers (both domestic and international) contributed to the successful transfer of power during the 1991 Zambian elections, but that observers struggled with their own political legitimacy vis-à-vis domestic audiences.²⁵ In El Salvador, Montgomery argues that the Organization of American State's hasty judgment of the 1991 elections and their failure to criticize blatant attempts to manipulate the election contributed to the government's decision not to reform obvious problems prior to the 1994 elections.²⁶ These and other similar studies offer a

wealth of information on elections in dozens of countries throughout the developing world, and set the foundation for this study.²⁷

A shared weakness of existing cross-national and case-study research on election monitoring, including some of my own research, is that such studies have difficulty attributing causal effects to international observers. Numerous studies show that the presence of observers is correlated with a variety of positive and negative outcomes following elections, but they have difficulty comparing the outcomes of observed elections to a counterfactual world in which observers were not present. As Carothers has argued, the most significant potential effect of international observers is difficult to measure:

Out of fear of being caught by foreign observers, political authorities may abandon plans to rig elections. Of course, few foreign officials would readily acknowledge having had such plans, making it hard to measure precisely the deterrent effect of electoral observation.²⁸

Knowledge that international observers will be present at an election may prevent fraud from being attempted by political parties and candidates, or may deter other illegal or improper behaviors, although the nature of the decision to invite international observers and engage in illicit activity makes a cross-national causal test of this hypothesis exceedingly difficult.

The pre-election prevention of electoral irregularities, as highlighted by Carothers, is the ideal outcome for organizations interested in encouraging democratic elections. However, another potentially important effect of observers is also possible. On election day, because observers are physically present in polling stations, they may have direct and micro-level effects on the election day behavior of voters, parties, candidates, or election officials. To use the most obvious example, individuals engaged in ballot box stuffing may not wish to carry out their plans in the physical presence of international observers. Similarly, polling station officials may be more likely to follow official electoral regulations if they are internationally observed. The

mechanism behind this “observer effect” is that individuals frequently behave differently when they know they are being watched, particularly if they are aware that they are engaging in illegal or socially undesirable behavior. Thus, the behavior of voters, poll workers, or party operatives could change as a result of a visit by international election observers, and meaningful behavioral change at this micro-level should be reflected in votes cast on election day.

Although there are many potential effects of observers and international actors on the quality of the electoral process before, during, and after election day, I focus here on a subset of potential effects: whether international election observers have direct effects on election day behavior. By randomly assigning international observers for their election day observation, all other variables are held constant, and any difference between the group of observed and unobserved areas can be causally attributed to international observers.

For democracy promoters and for academics, this study is ideal for replication. Randomization of short-term observers could be adopted as standard practice by election monitoring missions, and overtime, the result would be a clearer understanding of the potential effects of observers on elections, the conditions under which they have desired effects on the quality of elections, and the most efficient design of election monitoring missions.

Random Assignment and the Effects of International Election Observers

I illustrate how experimental methods can be applied to democracy assistance with a field experimental test of whether international election observers influence the election day behavior of voters, political party representatives, or election officials as reflected in the pattern of votes cast on election day. If international observers do not influence behavior on election day, randomly selected areas of the country should be equivalent across all observable indicators. If observers do cause a significant change in behavior on election day, the areas of the country that

were randomly assigned to be observed should be significantly different from those that were not.

For the 2004 presidential elections in Indonesia, I had the opportunity to attempt random assignment of international observers for the Carter Center's election day deployment. To my knowledge, this was the first attempt of this type within international election observation, and one of the first attempts in the more general field of democracy promotion.²⁹ The case of Indonesia was selected because the opportunity to attempt random assignment of international observers was made available. The introduction of randomly assigned international observers had been met with some skepticism by other practitioners. Although international election observation missions regularly use randomization to assign international observers to vote-counting centers at the end of election day as part of a parallel vote tabulation,³⁰ random assignment of short term election observers during voting was thought unnecessary, logistically too difficult, or contrary to some of the other goals of election observation.³¹ Random assignment of election day monitors is not standard practice, even though it carries other advantages for election monitoring missions, such as providing a defined sample of polling stations from which to draw their summary conclusions. Current practices for observer deployment vary by election and by organization, but the most common deployment method is to allow individual observer teams to choose the polling stations that they visit within a given region after they have been deployed throughout the country. This system may create biases in the observations collected by observers, as individual teams may use different methods to select where they observe.

The 2004 presidential elections in Indonesia were the first direct presidential elections in the country's history. Legislative elections held in 1999 and in April of 2004 were widely considered successful given the country's size and its newly democratizing status.³² Prior to

these elections, the president was selected indirectly. The incumbent in the 2004 elections, Megawati Sukarnoputri, had been in office since her 2001 appointment by the People's Consultative Assembly. There were two rounds of the 2004 presidential election; this article focuses on the second round runoff between the incumbent candidate Megawati Sukarnoputri (commonly referred to as Megawati or Mega) and the leading challenger, Susilo Bambang Yudhoyono (commonly referred to as SBY).

Expectations were high leading up to the 2004 elections, which were viewed as a crucial step in Indonesia's democratization.³³ Many believed that the elections were likely to go well, and the most common concerns in advance of election day pertained to logistical factors and the administration of an election in such a large and diverse country.³⁴ However, because of the scope of the election reforms leading to the 2004 elections and the recent transition to democratic institutions, some observers worried that the election could deteriorate into violence or fraud.³⁵

Prior to the election, there were reports of "money politics" and other forms of intimidation, complaints related to restrictions placed on domestic election observers, as well as violations of laws restricting campaign activity.³⁶ Despite these complaints, the overall environment leading up to the presidential elections was guardedly optimistic, and observers hoped that the election would be carried out peacefully. Thus, in the case of Indonesia in 2004, the anticipated effect that international observers could have on election day behavior was moderated by the expectation that the election would be relatively clean.

Clear-cut cases of blatant election day fraud would have made a more straightforward baseline study of whether election observers improve election quality. Theoretically speaking, Indonesia was a more complicated case. Although many experts in Indonesia politics were anxious in advance of the election, blatant election day fraud—such as ballot box stuffing—was

not expected.³⁷ In countries that experience widespread election day manipulation, the party of the incumbent government is frequently the primary sponsor, and other research has show that observers can deter blatant election day manipulation (Author). However, in Indonesia's 2004 election, the incumbent had never stood for a direct presidential election, and did not have a reputation for carrying out widespread election day manipulation. Additionally, going in to the second round runoff, Megawati had already lost the first round of the election to SBY, and was not expected to win. Thus, in designing this study, it was not clear in advance of the election which candidate would be more likely to benefit from the presence of observers.

Logistics of Implementation

In part due to the size of the country and the application of a randomized field experimental methods to untested circumstances, there were five logistical challenges in the implementation of this experiment, all of which were ultimately possible to mediate. I detail them here to make evaluation of the experiment more transparent and to facilitate future applications of field experimental methods to democracy promotion programs. Additionally, the discussion of logistical challenges illustrates that the conditions for this experiment were less than ideal, yet the study was still conducted successfully. This is a point that is often underemphasized when students and practitioners are trained in field experimental methods. Like in many other areas of teaching methods of research or evaluation, the emphasis tends to be on the ideal conditions for applying the methodology rather than how to address potential deviations from that ideal. This experiment illustrates that it is possible to carry out successful randomized evaluations of democracy promotion programs under somewhat adverse conditions while highlighting the risks and limitations inherent in making such tradeoffs.

Indonesia is one of the largest and most geographically diverse election-holding countries in the world, and in terms of votes cast, the presidential election was the largest single day election ever held in the world.³⁸ Muhammad Qodari describes the immensity of the challenge for the Indonesian election commission, or *KPU*, leading into 2004 series of elections:

The *KPU* had to print and distribute not only a unique identify card for each of [the] nearly 150 million voters, but also 660 million ballot papers... For those ballots to be of any use the *KPU* had to provide for the acquisition, transport, positioning, handling, and supervision in 32 provinces and 440 districts of 580,000 voting stations, 2.9 million voting booths, 2.3 million ballot boxes, and 1.2 million bottles of ink with which officials could mark voters' fingers in order to prevent multiple voting. Consider that all this had to be accomplished in a developing, still somewhat infrastructure-poor country that consists of more than 12,000 islands spread across... about 2% of the Earth's entire surface area—and one gets some sense of the awesome challenges involved.³⁹

In a much smaller country with somewhat better infrastructure and fewer islands, the ideal experimental design might have randomized the assignment of observers across the entire country. The first logistical challenge in the Indonesian case was that many areas of the country were not accessible to international observers on election day, and therefore random assignment could not be attempted across the entire population of polling stations. This issue was further exacerbated by the limited number of Carter Center observers participating in the randomization. Rather than randomize across the entire population, it was conducted within a significantly smaller group of pre-selected districts, or “blocks,” as described below.

The second logistical challenge was that there was no complete list of polling stations available from the central government. This problem was addressed by randomizing at the village level. Within Indonesia, there are five levels of administrative divisions pertaining to elections. In addition to the provinces (*propinsi*) and districts (*kabupaten or kota*) listed in the above quote, the districts for the September 2004 runoff election were divided into 4,987 sub-districts (*kecamatan*), the sub-districts were divided into approximately 60,000 villages or

neighborhoods (*kelurahan* or *desa*), and the villages and neighborhoods were divided into approximately 580,000 polling stations (*TPS*). In some cases the ideal research design would be to randomly assign observers to polling stations within each district where international observers were sent. However, in the case of Indonesia, even if there had been a complete list of polling stations, observers would have had a difficult time locating them. Many were set up outdoors at locations without physical addresses such as community badminton courts, in the middle of streets, on sidewalks, in empty lots or field, etc. Additionally, many polling stations were set up adjacent to each other, particularly in urban areas, so it would have been difficult for observers to visit one polling station without making their appearance known to adjacent polling stations, and the issue is further complicated because data are not systematically available on the adjacency of polling stations.

The best option was to randomly assign observers to visit the next largest administrative divisions: *kelurahan* and *desa*. These administrative divisions can be understood as villages in non-urban areas or neighborhoods within cities. Within the areas included in this study, they can be as small as a few dozen voters with just one polling station, or as large as 60,000 voters.⁴⁰ Most villages/neighborhoods are identifiable on a local map, making it possible for the observer teams to find them. This made random assignment across units logistically possible, both because a complete list of villages and neighborhoods existed, and because international observer teams had a reasonable chance of being able to identify and locate the treatment units on election day.

Third, it was unclear prior to the election whether any disaggregated election results would be made available. Had the government failed to release disaggregated election results, the lack of data would have prevented most tests of whether observers influenced election day

behavior. This issue was not possible to address in advance, but the government indicated that it would release these results prior to the election, and ultimately did so. Although the public release of disaggregated election results has become more common in many countries, in many cases the results are only posted for a limited period of time, and this could be a relevant issue in future studies.⁴¹

Fourth, because election law mandated that each polling station have no more than 300 voters, a full-length election day was determined unnecessary, and polling stations were only open from 7:00 a.m. to 1:00 p.m. for the presidential election, significantly limiting the number of polling stations that an observer team could visit on election day. This challenge did not prevent the experiment from taking place, but reduced the number of villages visited by observers, and the overall size of the experiment.

Finally, because it was the first time that random assignment of international election observers had been attempted (to my knowledge), many of the challenges in applying this methodology to election observation had yet to be worked out and agreed upon by the interested parties, and were negotiated in a very short time frame. Although this explains some of the decision-making timeline, the cooperation and flexibility of the Carter Center staff and delegation made the project feasible.

Background and Experimental Design

The Carter Center's mission for the second round of the election consisted of 57 observers and 28 observer teams, 23 of which were asked to participate in the study. Ultimately, missing election results in some regions and one team's decision not to follow the experimental protocol reduced the number of teams in the study to 19.⁴² The long-term election observers and the Jakarta-based staff of the Carter Center selected areas of Indonesia (primarily *kabupaten* and

kota, or districts and cities) where the Carter Center would send election observers. This selection of districts was intended to place Carter Center observers throughout the country, but was constrained by logistical and safety issues.⁴³ In order for an area to be selected, it had to be accessible by car or aircraft within one day's travel time, and had to have basic accommodations for the observer team that were judged as sufficiently safe.⁴⁴ There was also some effort made to avoid extensive overlap with the European Union election observation mission (the largest observer mission in the country), as well as consideration for whether access was granted to areas of Indonesia where foreigners are frequently prohibited from traveling such as Banda Aceh, Ambon, and parts of Papua. For the participating teams, random assignment was applied within each district or pair of districts (*kabupaten* and *kota*) where Carter Center observers were deployed.

The lists of villages and neighborhoods assigned to each participating Carter Center team were generated within each pre-selected geographic area using systematic random sampling (also known as patterned sampling).⁴⁵ Randomization requires that every unit within a given block has an equal probability of being selected.⁴⁶ Once the random assignment was conducted for each pair of observers, the lists were not released to anyone outside of the Carter Center staff and the observers assigned to each block.

The unit of analysis in this study is the village/neighborhood. However, note that each village or neighborhood contained one or more polling stations. Upon arriving at randomly selected locations, observers visited between one and four polling stations within each village or neighborhood. Within a given neighborhood or village, they were necessarily limited to those polling stations that they could locate. Most teams were able to spend time the day before the election scouting the area and looking for signs that polling stations were being built. They were

instructed not to choose polling stations based on any substantive characteristics such as the number of voters, complaints about the polling station, known popularity of one candidate, or recommendations from local officials or police. Rather, if a team went to more than one polling station in a given village, they were instructed to go to every third or fifth polling station that they could locate.

The random assignment of observers to villages or neighborhoods within each block generates two groups: the treatment group (which was assigned to be observed) and the control group (which was assigned not to be observed). In theory, the randomization should produce two groups that are equivalent except that one group was assigned to be “treated” with international election observation. Although it is unlikely, it is possible that randomization produces groups of villages/neighborhoods that are different in important ways, and could potentially generate misleading results. Therefore, I also check the degree to which the two groups are similar based on a variable that could not have been affected by the presence of election day observers, and should therefore not be significantly different between groups. When abundant data about the experimental population is available, such a randomization check is straightforward. Because this was the first direct presidential election in Indonesian history, and the first election for which I have been able to collect disaggregated election results, little historical precedent existed on which to base predictions of vote patterns for the 2004 elections. Nevertheless, because voter registration took place before election day, the average number of registered voters between the treatment and control villages can be used as a variable for the randomization check, as there is no reason that this variable should be significantly different between the treatment and control groups. Table 1 presents the results of the randomization check. Across all 19 blocks, assignment to the treatment group is not consistently or significantly

related to the number of registered voters, as expected. When all blocks are pooled together, as shown in the last column of Table 1, assignment to the treatment group is unrelated to the number of registered voters, indicating that there is no significant difference in the average number of registered voters in the treatment and control groups.

[Table 1 about here]

Table 2 summarizes the areas observed by Carter Center observers at the village level. Out of all villages in the visited regions, Carter Center observers were assigned to visit 482 villages, 95 of which were actually visited.⁴⁷ Within these 95 assigned and visited villages, 147 individual polling stations were visited. Note that a small proportion of villages in the control group were visited.⁴⁸ The so called “failure to treat” is well known in field experiments, and is discussed in greater detail below.⁴⁹

[Table 2 about here]

Data and Results

In the second round of the 2004 presidential elections, Susilo Bambang Yudhoyono and his running mate Jusuf Kalla were the leading candidates, having won 34% of the votes cast in the first round in a five candidate field. The incumbent president, Megawati Sukarnoputri won 27% in the first round. The runoff was held on September 20, 2004. According to official results, SBY won the presidency with 60.6% of the vote.

Government-reported unofficial election results were recorded for the total number of votes cast for each candidate for all villages in the second round of the 2004 presidential election and the total number of registered voters.⁵⁰ The unofficial results were made public by the Indonesian *KPU* (the general elections commission) for most of the country. These aggregate

results were uploaded by regional election officials to a central government-run website, and should be viewed as “unofficial” or uncertified government-provided election results.

Table 3 presents aggregate summary statistics for the 1,822 village-level observations included in the study. These unofficial election results were downloaded, compiled, and merged with international observer data by the author. All comparisons only include districts where Carter Center observers were deployed, where they participated in the randomization, and where village level elections results were reported for the entire district.

[Table 3 about here]

As mentioned above, Carter Center observation teams did not visit all villages that were randomly assigned to the treatment group, leading to some “failure to treat.” This issue was anticipated to some degree, and is common in other similar field experiments when it is difficult to ensure high levels of compliance in the field, or where many relevant variables such as travel time or the precise location of treatment units are difficult to collect in advance of the treatment. Despite the use of the word failure, failure to treat does not threaten the validity of most field experiments, although it requires careful attention.

In this experiment, one may be tempted to simply compare the subset of villages actually visited by international observers with those that were not. This comparison is likely to yield biased estimates. Removing untreated villages from the treatment group into the control group makes the study observational rather than experimental and takes away the central advantages of randomization.

To further clarify this point, in Indonesia, it is plausible that some villages were more difficult for observers to locate than others and that this “findability” determined which villages in the treatment group were actually visited. It is possible that “findability” is also related to

voting behavior or support for particular candidates. Therefore, it cannot be assumed the determination of which villages were actually monitored was also random.

The most straightforward method of analyzing the results from the experiment is to compare the villages assigned to the treatment group to the villages assigned to the control group within each geographic area across which observers were assigned. In experimental jargon, this estimate is the “intent-to-treat” effect (or ITT effect) on the dependent variable within each block. There are several other methods that could be used to estimate the effect of observers, given randomization within blocks and variation in the treatment rates across blocks. As the central dependent variable of interest, I use the natural log of the total number of votes cast for Megawati in each of the 1,822 villages and neighborhoods included in the study. It is not possible to observe what would have happened in the villages visited by observers if they were not, in fact, monitored. Instead, randomization allows a comparison of two groups of villages that should be alike in all ways except that one group was assigned to be visited by international observers.⁵¹

The random assignment of units means that it is possible to estimate the average ITT effect without accounting for any other observed differences between villages, but regression allows the inclusion of covariates and serves to reduce the unexplained variance in votes cast for Megawati. I calculate the ITT effect using ordinary-least squares (OLS) regression. To restate, the central dependent variable is the performance of the incumbent candidate, measured as the natural log of the total number of votes cast for Megawati in each village. An additional independent variable measuring the total number of registered voters in the village (logged) is included in the model.⁵²

Table 4 presents the estimated effect of being assigned to the treatment group within each regional block, and a pooled estimate across all areas included in the study. Even given the relatively low rate of assigned villages that were actually visited (as shown in Table 2), assignment to the treatment group is associated with *improved* performance for Megawati in 15 out of the 19 blocks. The consistent direction of the effect across more than three fourths of the blocks is unlikely due to chance.⁵³

In the last column of Table 4 provides a pooled estimate with fixed effects for each experimental block. Note again that *Treatment Group* is a measure of the assignment of observers and their “intent” to treat the village, not the actual presence of observers on election day. Because *Treatment Group* is dichotomous and the dependent variable is logged, the coefficients represent the percent change in total votes cast for Megawati given that *Treatment Group* changes from zero to one and all else is held constant. In the estimate in Table 4 including all districts in the study, assignment to the treatment group caused a 6.5% positive change in the number of votes cast for Megawati. To put this number in context, the average number of votes cast for Megawati per village is 1,394, and assignment to the treatment group is associated with an average increase of about 91 votes for Megawati across all villages in the treatment group.⁵⁴ The same estimates were conducted on votes cast for SBY, and are included in the online appendix. There is no significant relationship between observer presence and the performance of the winning candidate, SBY, indicating that the increase in votes for Megawati did not come directly at the expense of SBY. Estimates using vote share for each candidate as the dependent variable rather than votes cast produce similar results, but give smaller villages equal weight as larger villages.⁵⁵

[Table 4 about here]

The ITT estimates of the effect of observers on Megawati's vote share are diluted by the low treatment rates, and an observer effect was detected despite the fact that many of the villages assigned to the treatment group were never observed. Yet because assignment to the treatment group made it more likely that a given village or neighborhood would be visited, it is possible to estimate what the average size of the effect of observers on only those villages and neighborhoods that were actually visited.⁵⁶ This method still utilizes the random assignment of observers, and treats the actual visits by observers to a village as a function of their assignment to the treatment group. The full table is confined to the (online) appendix. Like the estimates presented in Table 4, total registered voters (logged) are included as an independent variable. Across all geographic blocks, when accounting for the low treatment rates, the estimated effect of observers on the internationally observed villages is associated with a +32% change in votes cast for Megawati, which translates into an average increase of 446 votes per treated village.⁵⁷

Discussion

Overall, the results of this field experiment show that the incumbent candidate performed better and the challenger performed about the same in villages and neighborhoods assigned to be monitored by Carter Center observers. This result was not anticipated, and highlights a central advantage of using field experimental methods: the possibility that they can reveal effects that are not anticipated by scholars or practitioners.⁵⁸ Such a surprising result nevertheless requires some speculative explanation and analysis of the unique circumstances surrounding this election. Why might the presence of observers increase votes cast for Megawati, but not decrease votes cast for SBY? Why did observers influence what was widely viewed as a democratic election?

The reports of international observers, journalists, and analysts suggest several possible explanations. Although all major international observer organizations judged the observed

problems with the election to be insignificant, a number of irregularities were documented and described in the post-election reports of international observers. The most plausible explanation for this finding stems from the early closing of polling stations. The official election day was from 7:00 a.m. to 1:00 p.m., but after the first round of the presidential election, the *KPU* ruled that polling stations could close after 11:30 provided that all eligible voters had voted. If this rule was followed correctly, it should not have produced significant problems, and only those polling stations which reached 100% turnout should have closed early. Reports suggest, however, that a number of polling stations closed before all eligible voters had cast a ballot, and well before the earliest legal closing time of 11:30 (EU 2004, 58, Carter Center 2005, 63).

The presence of observers could have influenced the decision by election officials to close early by making it more likely that polling stations in visited areas would stay open until the mandated time. Additionally, during the course of their observation, many Carter Center observers announced or implied that they could return later in the day to observe the closing. If Megawati supporters were less likely to turn out to vote without being mobilized to do so by party representatives or election officials, correctly following the regulations surrounding the length of election day would have disproportionately benefited Megawati voters. Local party officials would have more time to mobilize voters, and poll-workers would have had greater incentive to prove that all voters had cast a ballot so that they could close early without violating electoral regulations. One potential explanation is therefore that non-observed villages were more likely to close before less motivated or reluctant voters had shown up, and were less likely to follow the electoral regulations about staying open until 1:00 p.m. or until all registered voters had cast a ballot.

Several additional pieces of evidence support the possibility that Megawati supporters were more reluctant to turn out and also suggest that she was not in control of the party or state machinery that would have been required to engage in widespread election day fraud. First, her party performed poorly in the April legislative elections and in the first round presidential elections. Second, in the weeks leading up to the run-off election, it was widely speculated in the media that she would lose, with public opinion polls from several organizations predicting support for SBY at about 60% and support for Megawati at around 29%.⁵⁹ Third, although Megawati had some incumbency advantages, including the ability to make public appearances throughout the country outside of the legal three-day campaign period, her support from several prominent parties was unstable. For example, Megawati was endorsed by the powerful Golkar party, which won the April 5th legislative elections, and which possessed well developed local party machinery that could have been used to mobilize the vote for Megawati. But several weeks before the election, national and local party leaders publicly split over the decision to endorse Megawati, and before the election analysts predicted that “Golkar will not be able to fully bring its formidable party machinery behind Megawati.”⁶⁰ Post-election polling revealed that the vast majority of Golkar voters who cast a ballot voted against their party’s endorsement and for SBY.⁶¹ Relative to incumbent presidential candidates in other countries, Megawati’s election day advantage was minimal.

If Megawati supporters were reluctant to turn out, she should perform better in those areas where turnout was higher. Although it is not conclusive evidence, scatter plots of votes cast for Megawati vs. turnout across all 1,822 villages included in the experiment (shown in the online appendix) illustrate that that Megawati does somewhat better in villages with higher turnout and SBY does worse, on average, in villages with higher turnout. These comparisons do

not prove that increasing turnout would have necessarily increased votes for Megawati, but they are consistent with the idea that Megawati's supporters were more reluctant to turn out, and that her performance increased when voter mobilization increased.

There are other less plausible potential explanations for Megawati's increased support within monitored villages. Reports from the Carter Center and the EU missions highlight numerous complaints of "money politics," including vote buying and the inappropriate use of government resources to support particular candidates. Few of these complaints were documented directly by observers, and there is little to suggest that vote-buying on election day was occurring. Of course, successful bribery and intimidation may be invisible to all but the participants. If intimidation was taking place in the second round in favor of SBY, it is possible that voters in monitored villages felt more confident in voting for Megawati when observers were present, although I have found little anecdotal support for this scenario.

Could observers from the Carter Center have caused extra support for Megawati? This explanation is similarly unlikely. Recall that international observers were mobile throughout election day, traveling from polling station to polling station. It is technically possible that those who were not inclined to vote on their own were drawn to the polling station because foreign observers visited. It remains a possibility that a visit by observers attracted additional voters to the polls, but it is not clear why this might have disproportionately influenced Megawati supporters.

The results presented in this article show a clear difference between observed and unobserved villages, but they are subject to interpretation. The most likely explanation for this finding, in my view, is that observers made polling station officials more likely to follow electoral regulations, and therefore caused visited polling stations to stay open later than they

would have if observers had not visited. Given that the election was expected to be relatively free of election day irregularities, the fact that any significant effect of observers was found is noteworthy. This result does not imply election fraud. If widespread election fraud by one candidate had taken place, and this fraud was deterred by observers, the cheating candidate should have performed worse in areas that were observed. Even though Megawati benefited from observers, the results do not show that SBY performed significantly worse when observers were present, as would be expected if observers reduced ballot box stuffing or other forms of direct election fraud. Rather, I argue that election officials were more likely to follow the letter of the election law pertaining to closing time after having been visited by international observers.

The Carter Center mission concluded that “voters were able to exercise their democratic rights in a peaceful atmosphere and without significant hindrance” (Carter Center 2005).⁶² The results presented here do not contradict this conclusion. Even so, this article demonstrates that international observers had measurable effects on election day behavior, causing localized improvement in the performance of the losing incumbent presidential candidate. This specific result is somewhat idiosyncratic, but the fact that it was unanticipated highlights one of the central advantages of field experiments in that they allow researchers to uncover effects of interventions even if they are not anticipated. More generally, the study illustrates the potential use of field experimental methods to evaluate the effects—anticipated or not—of democracy assistance programs like international election monitoring.

Conclusion

This article has outlined how random assignment of international election observers can be used to study whether international actors influence electoral behavior, and presented the randomization of election observers as an example of how individual components of democracy

promotion programs can be evaluated. In the case of the second round of the 2004 presidential election in Indonesia, international observers were randomly assigned to villages within defined geographic areas. The evidence suggests that the presence of observers caused increases in total votes cast for the incumbent, Megawati Sukarnoputri, who went on to lose the election and peacefully transfer power to her competitor. These results suggest that even in a relatively clean election, observers can change election day behavior in a manner that can disproportionately benefit some candidates.

I also discussed this experiment in the larger context of using field experimental methods to evaluate democracy promotion. This experiment –and other like it—are ideal for replication in other settings, and similar field experimental methods should be applied to advance our understanding of the effects of election observation and of other democracy promotion activities. Across many other areas of democracy promotion, the potential exists for increased learning about the causal effects of specific democracy promotion programs. Potentially foreshadowing the future of this method in democracy promotion, the field of development economics has recently experienced a dramatic increase in the use of field experimentation, and scholars and policy-makers in some parts of the field are now working together in long-term cooperative relationships aimed at what Banarjee and Duflo call an “iterated process of policy learning” whereby field experimentation is deployed as a recurrent element of program evaluation and the lessons learned from previous studies help inform future policy making and the design of additional field experiments. In this model, field experiments are not a one-shot activity, but built in to a long term plan to understand the conditions under which various programs are effective. There are many criticisms of field experimentation, including that the trend will devolve into “only” program evaluation, that experiments are unethical and interventionist, or that they are

unlikely to answer interesting and important questions. However, democracy promotion is at least one issue area that has enormous potential for mutually beneficial learning and cooperation between academics and practitioners. In response to a similar debate in development economics, Banarjee and Duflo defend the use of field experiments and their relationship to policy-making:

To be interesting, experiments need to be ambitious, and need to be informed by theory. This is also, conveniently, where they are likely to be the most useful for policymakers. Our view is that economists' insights can and should guide policy-making... They are sometimes well placed to propose or identify programs that are likely to make big differences. Perhaps even more importantly, they are often in a position to midwife the process of policy discovery, based on the interplay of theory and experimental research.⁶³

I have intentionally presented a very optimistic view of the potential use of field experiments in democracy promotion. Increased cooperation between policy makers and academics has long been a goal, and the potential payoff could be enormous: combined with existing research methods, a long-term cooperative relationship would ideally play a central role in revealing the conditions under which various democracy promotion programs have their intended effects, which types of democracy promotion are most efficient, the conditions under which specific programs are most likely to have positive or negative effects, and whether such interventions have unintended effects. Although these payoffs are far from certain, and debates over the value of field experiments will certainly continue, the potential benefits are great.

Table 1
Logistic Regression of Assigned-to-Treatment Group on Registered Voters

<i>Block Number</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
<i>Registered Voters (in 1000s)</i>	-0.103 (0.215)	0.028 (0.019)	-0.001 (0.110)	-0.030 (0.102)	0.104 (0.136)	0.124 (0.074)	-0.078 (0.144)	0.175 (0.128)	0.006 (0.032)	-0.004 (0.043)	
Constant	-1.125* (0.471)	-1.704** (0.333)	-0.072 (1.231)	-1.173** (0.355)	-1.500** (0.451)	-1.645** (0.353)	0.345 (1.086)	-1.515* (0.674)	-1.495** (0.379)	-1.265** (0.202)	
Obs.	90	163	23	186	117	136	45	46	156	243	
LR $\chi^2(1)$	0.24	1.97	0.00	0.09	0.57	2.81	0.30	1.93	0.04	0.01	
Prob > χ^2	0.628	0.160	0.990	0.770	0.451	0.093	0.587	0.165	0.851	0.929	
<i>Block Number</i>	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	Pooled Estimate
<i>Registered Voters (in 1000s)</i>	-0.021 (0.054)	0.004 (0.059)	-0.029 (0.075)	-0.058 (0.048)	0.160* (0.078)	-0.007 (0.060)	-0.097 (0.223)	0.170 (0.235)	-0.230 (0.136)	0.016 (0.048)	0.011 (0.012)
Constant	-1.231 (0.674)	-1.415** (0.273)	-1.326** (0.513)	0.339 (0.801)	-2.304** (0.524)	-1.363* (0.576)	-1.197* (0.530)	-1.645** (0.586)	-0.015 (0.496)	-1.527** (0.346)	-1.34 (0.259)
Obs.	42	343	31	24	103	103	60	52	56	146	2165
LR $\chi^2(1)$	0.15	0.01	0.16	1.57	4.19	0.01	0.19	0.474	3.21	0.11	32.33
Prob > χ^2	0.697	0.939	0.685	0.210	0.041	0.908	0.659	0.010	0.073	0.746	0.040

Notes: Model 21 includes dummy variables for each block (not reported). Standard errors in parentheses. *significant at 5%, **significant at 1%.

Table 2
Carter Center Observation Coverage

Block	Province	District or City	Number of Villages in Treatment Group	Number of Villages in Control Group	Number Villages Observed	Villages Visited in Control Group	Total Voters
1	Nanggroe Aceh Darussalam	Kota Banda Aceh	19	71	3	1	173,265
2	East Java	Kota Surabaya	34	129	3	0	2,078,486
3	Nusa Tenggara Barat	Kota Mataram	11	12	4	0	241,483
4	East Java	Sampang	41	145	5	0	569,216
5	Bali	Tabanan	27	90	6	0	325,701
6	East Java	Situbondo	32	104	10	0	488,633
7	DI Yogyakarta	Kota Yogyakarta	20	25	4	0	327,873
8	East Java	Kota Kediri	15	31	11	0	200,137
9	North Sumatera	Kota Medan	30	126	5	2	1,525,526
10	Riau	Kampar, Kota Pekanbaru Baru	53	190	5	0	740,924
11	East Kalimantan	Kota Samarinda	8	34	4	2	453,693
12	West Java	Cianjur	68	275	1	3	1,378,863
13	Central Kalimantan	Kota Palangka Raya	6	25	1	4	123,596
14	West Kalimantan	Kota Pontianak	9	15	4	3	371,780
15	West Sumatra	Kota Padang	20	83	4	1	525,422
16	South Sumatra	Palembang	20	83	11	1	906,169
17	North Sulawesi	Kota Bitung	12	48	3	4	120,637
18	North Maluku	Kota Ternate	11	41	5	1	95,771
19	Maluku	Kota Ambon	18	38	4	1	192,097
20	South Sulawesi	Kota Makassar	28	118	2	2	812,977
TOTAL (with Block 12)			482	1683	95	23	11,652,249
TOTAL (without Block 12)			414	1408	94	22	10,273,386

Table 3
Summary Statistics for All Available Village-Level Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Observed	1822	0.064	0.224	0	1
Sample	1822	0.227	0.419	0	1
Megawati (Total Votes)	1822	1394	1662	1	22494
SBY (Total Votes)	1822	2486	2538	5	19618
Overall Turnout	1822	0.715	0.117	0.1	1
Ballots Received	1822	5645	5826	35	59567
Valid Ballots	1822	3880	3848	6	42112
Invalid Ballots	1822	83	131	0	1582
Extra Ballots	1822	6	26	0	345
Damaged Ballots	1822	43	276	0	5923
Ballots Not Used	1822	1645	2071	0	16612
Total Registered Voters	1822	5639	5821	35	59567

Table 4
OLS: Estimated Effects of Intent to Treat on Total Votes for Megawati (ln)

<i>Block Number</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Treatment Group</i>	0.14 (0.11)	0.02 (0.05)	0.07 (0.15)	0.24* (0.10)	0.02 (0.06)	0.15* (0.07)	0.11 (0.06)	0.11 (0.06)	-0.07 (0.09)	0.03 (0.09)
<i>Reg. Voters (ln)</i>	0.97** (0.06)	1.04** (0.03)	0.88** (0.19)	1.36** (0.08)	0.85** (0.05)	0.94** (0.05)	0.90** (0.10)	0.91** (0.05)	1.04** (0.03)	0.86** (0.04)
Constant	-2.16** (0.42)	-1.65** (0.26)	-0.69 (1.78)	-4.34** (0.60)	0.68 (0.40)	-0.81* (0.41)	-0.50 (0.89)	-0.29 (0.39)	-2.00** (0.29)	-0.88** (0.28)
Observations	90	163	23	186	117	136	45	46	156	243
R-squared	0.77	0.90	0.51	0.64	0.71	0.74	0.66	0.90	0.87	0.70

<i>Block Number</i>	(11)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	Pooled Estimate
<i>Treatment Group</i>	0.06 (0.10)	0.19 (0.21)	-0.11 (0.15)	0.03 (0.10)	-0.04 (0.04)	-0.14 (0.07)	0.19 (0.17)	0.04 (0.18)	0.16 (0.15)	0.065* (0.026)
<i>Reg. Voters (ln)</i>	0.87** (0.05)	0.87** (0.05)	1.14** (0.10)	0.80** (0.06)	0.89** (0.03)	0.83** (0.03)	0.51** (0.09)	0.99** (0.08)	0.92** (0.07)	0.94** (0.014)
Constant	-0.52 (0.45)	-0.42 (0.36)	-2.98** (0.94)	-0.86 (0.52)	-0.08 (0.25)	0.28 (0.25)	1.69* (0.64)	-0.95 (0.67)	-1.80** (0.55)	-1.95** (0.11)
Observations	42	31	24	103	103	60	52	56	146	1822
R-squared	0.89	0.91	0.88	0.63	0.91	0.92	0.42	0.73	0.58	0.87

Note: Pooled estimate includes block fixed effects. Standard errors in parentheses. * significant at 5%; ** significant at 1%

Notes

- ¹ Election monitoring and election observation are used interchangeably in this article.
- ² Bjornlund 2004; Huntington 1991; Hyde 2006; Kelley 2008; Rich 2001.
- ³ Burnell 2000; Carothers 2004; Smith 1994; National Research Council 2008.
- ⁴ Cox, Ikenberry, and Inoguchi 2000; Gillespie and Youngs 2002; Smith 1994; Newman and Rich 2004.
- ⁵ Guilhot 2005; Robinson 1996.
- ⁶ McFaul 2004; Rich 2001; Levitsky and Way 2005; Bjornlund 2004; Carothers 1999; Pevehouse 2005; Donno 2008; Cooper and Legler 2006; Legler, Lean, and Boniface 2007.
- ⁷ Finkel et al. 2007.
- ⁸ Barnett and Finnemore 1999.
- ⁹ 2007: 414.
- ¹⁰ Knack 2004; Djankov, Montalvo, and Reynal-Querol 2008; Wright 2009; Nielsen and Nielson 2008; Kalyvitis and Vlachaki 2008.
- ¹¹ Brown 1998; Ottaway 1998.
- ¹² Cooper and Legler 2006; Levitsky and Way 2005.
- ¹³ National Research Council 2008: 134.
- ¹⁴ Millennium Challenge Corporation 2009.
- ¹⁵ Bjornlund 2004.
- ¹⁶ Bjornlund 2001; Carothers 1999.
- ¹⁷ Burnell 2000; Gleditsch 2002; Pevehouse 2002; Pevehouse 2005; Whitehead 1996.
- ¹⁸ Gourevitch 1978.
- ¹⁹ Hyde 2007; Middlebrook 1998.
- ²⁰ OSCE/ODIHR 2005: 2.
- ²¹ Carothers 1997; Soremekun 1999.
- ²² Brown 2005; Geisler 1993.
- ²³ Pastor 1998: 155.
- ²⁴ Abbink and Hesselning 2000; Carothers 1997; Pastor 1998.
- ²⁵ 1992.
- ²⁶ 1998.
- ²⁷ See, for example, Bjornlund 2004; Beigbeder 1994; Bratton 1998; Geisler 1993; Kumar 1998; Laakso 2002; Matlosa 2002; McCoy 1998; Middlebrook 1998; Pastor 1998; Booth 1998; Nwankwo 1999; Oquaye 1995; Orozco 2002; Scranton 1998. On election manipulation more generally see Lehoucq 2003; Lehoucq and Molina 2002; Schedler 2002; Schedler 2006.
- ²⁸ 1997: 18.
- ²⁹ Since that time, randomized assignment of international observers has been conducted by The Carter Center in Nicaragua (2006), by a Yale University student delegation participating in a US Embassy mission in Mauritania (2007), and by NDI in the 2006 Palestinian elections.
- ³⁰ The parallel vote tabulation, or quick count, provides an independent measure of the election results, within a margin of error, and is traditionally more reliable than exit polling. Observers (domestic or international) are assigned to a random sample of polling stations to directly observe the counting process. They call in the tallies from the vote count, and because the sample is random, quick counts typically provide very accurate estimations of the election results, and thus guard against manipulation during the counting process. See Estok, Nevitte, and Cowan 2002.
- ³¹ For example, one strategy for election monitoring is to send observers to the areas that are expected to have problems, or to send observers to areas that would ‘benefit from seeing an international presence’. These strategies create clear bias in the content of election day observations, but are perceived as politically important. (Personal conversations between the author and international election observation professionals from NDI, the EU, the OSCE/ODIHR, and The Carter Center.) Of course, it would be possible to randomize within regions that are expected to have problems in order to alleviate this concern.
- ³² Qodari 2005.
- ³³ Emmerson 2004; Wanandi 2004; Liddle and Mujani 2005; Qodari 2005.
- ³⁴ Qodari 2005.
- ³⁵ Emmerson 2004; Carter Center 2005; European Union 2004.
- ³⁶ Aspinall 2005; Carter Center 2005; European Union 2004.

³⁷ Emmerson 2004.

³⁸ Qodari 2005.

³⁹ 2005: 77.

⁴⁰ Of the 1822 units included in the study, the average size of a *desa/kelurahan* is 5638 registered voters. The smallest unit is 35 registered voters, and the largest is 59,567 registered voters. Data from the KPU did not distinguish between *desa* and *kelurahan*.

⁴¹ A permanent collection of disaggregated election results across all countries which release them would be an enormous resource for academics and practitioners.

⁴² Unfortunately, data were incomplete for three of the districts where teams from the Carter Center were deployed in the second round: Mimika, Kupang and Manokwari. These regions (and the three corresponding Carter Center teams) were dropped from the analysis. Additionally, there is one block in particular (Block 12, in Cianjur) where the experimental protocol was not followed by the team of monitors assigned to that region. There was no experiment to speak of in Block 12, with monitors going only to less than 2% of both the treatment and control groups. It is also a block with an unusually large number of villages, representing a significant portion of the “failure to treat” villages. The reason for the failure of implementation in this block was that the team of monitors assigned there did not attempt to comply with the assigned list of villages, a decision that was not influenced by the characteristics of the block (see Nickerson 2005). Although I present the summary data for this block in Table 2, I exclude it from the remainder of the analysis. This issue is discussed in greater detail in a separate article (Author).

⁴³ This selection of districts and sub-districts is non-random, and the experiment cannot be interpreted as nationally representative.

⁴⁴ Security concerns are relatively standard on election observation missions, but were heightened in Indonesia because of recent Western-targeted bombings of hotels and the Australian embassy.

⁴⁵ For a given block (city or district) to which a Carter Center team was assigned, a complete list of villages and neighborhoods was compiled. The total number of units within each block, or N_i for all i blocks, was sorted by an identification number that roughly identified the units geographically, but was not otherwise organized in any systematic pattern. For each block, a target number of randomly selected units, n_i , was produced in negotiation with regional experts and the Carter Center staff, and for logistical reasons allowed a greater proportion of selected units within some blocks. Given n_i , every k th unit was selected, with $k = N_i / n_i$ for all i blocks. The first village chosen in the skipping pattern was selected arbitrarily from all villages within the block. The arbitrary selection was made by the author. It was not strictly random (say, by roll of dice) but the first selected village to start the skipping pattern was selected without attention to location.

⁴⁶ An alternative method for randomization would have been to use a computer program to generate a list of random numbers that correspond with each polling station, sort by the random numbers, and select the appropriate number of polling stations from the top of the list.

⁴⁷ The low treatment rate is due primarily to the nature of the negotiations with the Carter Center and the fact that this was the first study of its kind. There was a large degree of uncertainty about the number of villages that each team or pair of teams could visit on election day given the quality of roads, population density, and other logistical factors. The delegation leadership agreed to randomize most of their deployment plan on the condition that observers would be assigned more units than they were realistically expected to be able to visit, such that no team would run out of assigned treatment units. In practice, this generated low treatment rates, and is similar to voter mobilization experiments in which canvassers are assigned to treat voters at many more households than they can actually treat, as a number of unknown variables influence which households the canvassers will actually be able to contact (Arceneaux, Gerber, and Green 2006; Gerber and Green 2000).

⁴⁸ For some teams, visiting control group villages or neighborhoods was accidental and resulted from visiting polling stations near the border between urban neighborhoods. Other teams encountered logistical (usually transportation related) problems that caused them to choose to visit villages outside of their assigned list. This information is only available anecdotally, and was not coded in the dataset.

⁴⁹ I do not attempt a complete summary of this literature here. For an overview within the context of political science, see Arceneaux, Gerber, and Green 2006; Gerber and Green 2000; and Nickerson 2005.

⁵⁰ Data were downloaded from the KPU website, <http://tnp.kpu.go.id/>.

⁵¹ Here, the estimated ITT effect of observers on vote share within each geographic block (i) is the average difference in incumbent performance in villages between treatment and control groups. Expressing this

logic mathematically, if the average number of votes cast for Megawati within each block is represented across the treatment group as Y_i^T and in the control group as Y_i^C , then the parameter of interest for all i blocks is simply: $ITT_i = Y_i^T - Y_i^C$

⁵² This basic model used to estimate the intent to treat effect can therefore be expressed as:

$\log(Y_j) = a + b_1 T_j + b_2 \log(X_j) + \mu$, where Y is a continuous variable representing the total votes cast for Megawati in village j , $T_j=1$ if the village was assigned to the treatment group, X is a variable representing the total number of registered voters in the village, and μ represents unobserved causes of votes for Megawati.

⁵³ If improved performance was as likely as not, the binomial probability that 15 of 19 outcomes would be positive is 0.0096 (one-tailed test).

⁵⁴ Overall turnout (votes cast/registered voters) is also higher in monitored villages, although the effects is smaller and only marginally

⁵⁵ Additionally, I estimated all models with a variable indicating the presence of EU observers. EU observers were not randomly assigned. Out of the 1,822 villages included in this study, EU observers visited 61. Of these 61 villages, 4 were in the treatment group and also visited by Carter Center observers, and 6 were in the assigned treatment group but not visited by Carter Center observers. The inclusion of this variable has minimal influence on the sign and significance of the (randomized) Carter Center observation variable.

⁵⁶ In order to account for treatment rates, and following previous applications in field experiments, I use instrumental variable techniques to estimate the magnitude of the treatment effect, also known as the “Average Treatment Effect” or ATE (Angrist, Imbens, and Rubin 1996; Gerber and Green 2000). Very generally, this estimate can be understood as the ITT effect divided by the actual treatment rates within each block, and can be estimated using instrumental variables regression, as presented in Appendix Table 5. Using two-stage least-squares regression (2SLS), for an instrument to be valid, it must be correlated with the actual treatment (or the endogenous variable) but not correlated with the error term in the model. Assignment to the treatment group of villages within a region is random, and there is therefore no reason that *Treatment Group* should be correlated with the error term. Actual treatment, or being visited by international observers, is a function of a village being assigned to the treatment group. When the actual visit by observers to a given village is used as an explanatory variable, assignment to the treatment group satisfies the conditions for a valid instrument. Because of the diverse readership of *Perspectives*, I have excluded much of the technical discussion to this footnote and provide the results in an online appendix to this article.

⁵⁷ Note that both Tables 4 and 5 present pooled estimates in which all blocks are combined. In a separate paper I discuss the methodological decisions relating to this type of analysis using the Indonesian experiment as a case, and explore the consequences of various modeling decisions. The paper is available at <http://>.

⁵⁸ Banerjee and Duflo 2008.

⁵⁹ “Indonesia’s Megawati Heading for Defeat, Two Polls Show,” *Associated Press Worldstream*, September 15, 2004.

⁶⁰ “What Lies Ahead After Indonesia’s Election,” *United Press International*, September 14, 2004;

“Golkar Party Leaders Split as Internal Rift Deepens,” *The Jakarta Post*, September 1, 2004.

⁶¹ Liddle and Mujani 2005.

⁶² See also the EU final report (European Union 2004) and the JAMPPI final report (domestic observers).

⁶³ Banerjee and Duflo 2008; cf. Heckman 1991.

Works Cited

- Abbink, Jon, and Gerti Hesselning, eds. 2000. *Election Observation and Democratization in Africa*. Houndmills, England: Macmillan Press.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444-455.
- Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment." *Political Analysis* 14 (1): 37-62.
- Aspinall, Edward. "Elections and the normalization of politics in Indonesia." *South East Asia Research* 13: 117-156.
- Banerjee, Abhijit V., and Esther Duflo. 2008. "The Experimental Approach to Development Economics." *CEPR Discussion Paper No. DP7037*.
- Barnett, Michael N, and Martha Finnemore. 1999. "The Politics, Power, and Pathologies of International Organizations." *International Organization* 53 (4): 699-732.
- Beigbeder, Yves. 1994. *International Monitoring of Plebiscites, Referenda and National Elections: Self-Determination and Transition to Democracy*. Martinus Nijhoff Publishers.
- Bjornlund, Eric. 2004. *Beyond Free and Fair: Monitoring Elections and Building Democracy*. Washington, D.C: Woodrow Wilson Center Press.
- Bjornlund, Eric. "Democracy Inc." 2001. *The Wilson Quarterly*.
<http://www.democracyinternational.com/downloads/01_sum_wq.pdf>.
- Bjornlund, Eric, Michael Bratton, and Clark Gibson. 1992. "Observing Multiparty Elections in Africa: Lesson from Zambia." *African Affairs*. 91(364): 405-432.
- Booth, John A. "Electoral observation and democratic transition in Nicaragua." In *Electoral Observation and Democratic Transitions in Latin America*, edited by Kevin J. Middlebrook, 187-209, 1998.
- Bratton, Michael. "Second Elections in Africa." *Journal of Democracy* 9 (3) (1998): 51-66.
- Brown, Frederick Z. 1998. "Cambodia's Rocky Venture in Democracy." In *Postconflict Elections, Democratization, and International Assistance*, edited by Krishna Kumar. Lynne Rienner Publishers.
- Brown, Stephen. 2005. "Foreign Aid and Democracy Promotion: Lessons from Africa." *The European Journal of Development Research*. 17(2):179-198.
- Burnell, Peter J. 2000. *Democracy Assistance: International Co-Operation for Democratization*. London: Frank Cass.
- Carothers, Thomas. 1999. *Aiding Democracy Abroad: The Learning Curve*. Washington DC: Carnegie Endowment for International Peace.
- Carothers, Thomas. 2004. *Critical Mission: Essays on Democracy Promotion*. Washington DC: Carnegie Endowment for International Peace.
- Carothers, Thomas. 1997. "The Observers Observed." *Journal of Democracy*. 8,3: 17-31.
- Carter Center, The. 2005. *The Carter Center 2004 Indonesia Election Report*. Atlanta: The Carter Center. <<http://www.cartercenter.org/documents/2161.pdf>>.
- Cooper, Andrew F., and Thomas Legler. 2006. *Intervention Without Intervening?: The OAS Defense and Promotion of Democracy in the Americas*. Palgrave Macmillan.

- Cox, Michael, G. John Ikenberry, and Takash Inoguchi, eds. 2000. *American Democracy Promotion: Impulses, Strategies, and Impacts*. Oxford: Oxford University Press.
- Djankov, Simeon, Jose Montalvo, and Marta Reynal-Querol. 2008. "The Curse of Aid.," *Journal of Economic Growth*. 13(3):169-194.
- Donno, Daniela. 2008. *Defending Democratic Norms: Regional Intergovernmental Organizations, Domestic Opposition and Democratic Change*. PhD Dissertation, Yale University.
- Emmerson, Donald K. 2004. "A Year of Voting Dangerously?" *Journal of Democracy* 15 (1): 94-108.
- Estok, Melissa, Neil Nevitte, and Glenn Cowan. 2002. *The Quick Count and Election Observation: An NDI Handbook for Civic Organizations and Political Parties*. National Democratic Institute for International Affairs.
- European Union. *European Union Election Observation Mission to Indonesia, 2004*.
- Finkel, Steven E., Andrew Green, Anibal Perez-Linan, and C. Neal Tate. 2007. "Cross-National Research on USAID's Democracy and Governance Programs - Codebook (Phase II)."
- Finkel, Steven E., Aníbel Pérez-Liñán, Mitchell A. Seligson, and Dinorah Azpuru. 2007. "Effects of US Foreign Assistance on Democracy Building, 1990-2003." *World Politics* 59(3):404-440.
- Geisler, Gisela. 1993. "Fair? What Has Fairness Got to Do with It? Vagaries of Election Observations and Democratic Standards." *The Journal of Modern African Studies* 31 (4): 613-637.
- Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (3): 653-663.
- Gillespie, Richard, and Richard Youngs. 2002. "Themes in European Democracy Promotion." *Democratization* 9 (1): 1.
- Gleditsch, Kristian Skrede. 2002. *All International Politics is Local: The Diffusion of Conflict, Integration*. Ann Arbor: University of Michigan Press.
- Gourevitch, Peter. 1978. "The Second Image Reversed: The International Sources of Domestic Politics." *International Organization* 32 (4): 881-912.
- Guilhot, Nicolas. 2005. *The Democracy Makers: Human Rights and the Politics of Global Order*. New York: Columbia University Press.
- Heckman, James J. 1991. "Randomization and Social Policy Evaluation." *National Bureau of Economic Research Technical Working Paper Series No. 107*. Available at <<http://www.nber.org/papers/t0107>>.
- Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. University of Oklahoma Press.
- Hyde, Susan D. 2006. *Observing Norms: Explaining the Causes and Consequences of Internationally Monitored Elections*. PhD Dissertation, University of California, San Diego. <<http://proquest.umi.com/>>.
- Hyde, Susan D. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 60 (1): 37-63.
- Kalyvitis, Sarantis C., and Irene Vlachaki. 2008. "Democratic Aid and the Democratization of Recipients." SSRN eLibrary (October 6, 2008). <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=888262>.

- Kelley, Judith. 2008. "Assessing the Complex Evolution of Norms: The Rise of International Election Monitoring." *International Organization* 62 (02): 221-255.
- Knack, Stephen. 2004. "Does Foreign Aid Promote Democracy?." *International Studies Quarterly* 48 (1): 251-266.
- Kumar, Krishna. 1998. *Postconflict Elections, Democratization, and International Assistance*. Lynne Rienner Publishers.
- Laakso, Lissa. 2002. "The politics of international election observation: the case of Zimbabwe in 2000." *Journal of Modern African Studies* 40 (03): 437-464.
- Legler, Thomas, Sharon F. Lean, and Dexter S. Boniface, eds. 2007. *Promoting Democracy in the Americas*. The Johns Hopkins University Press.
- Lehoucq, Fabrice, and Ivan Molina. 2002. *Stuffing the Ballot Box Fraud, Electoral Reform, and Democratization in Costa Rica*. New York: Cambridge University Press.
- Lehoucq, Fabrice . 2003. "Electoral Fraud: Causes, Types, and Consequences." *Annual Review of Political Science*. 6: 233-256.
- Levitsky, Steven., and Lucan Way. 2005. "International Linkage and Democratization." *Journal of Democracy* 16 (3): 20-34.
- Liddle, R. William, and Saiful Mujani. 2005. "Indonesia in 2004: The Rise of Susilo Bambang Yudhoyono." *Asian Survey* 45(1): 119-126.
- Matlosa, Khabele. 2002. "Election monitoring and observation in Zimbabwe: Hegemony versus sovereignty ." *African Journal of Political Science* 7 (1): 129-154.
- McCoy, Jennifer. 1998. "Monitoring and Mediating Elections during Latin American Democratization." In *Electoral Observation and Democratic Transitions in Latin America*, edited by Kevin J. Middlebrook, 53-90.
- McFaul, Michael. 2004. "Democracy Promotion as a World Value." *The Washington Quarterly* 28 (1): 147-163.
- Middlebrook, Kevin J. 1998. *Electoral Observation and Democratic Transitions in Latin America*. La Jolla: University of California.
- Millennium Challenge Corporation. 2009. "Impact Evaluation at MCC," Available at: <<http://www.mcc.gov/programs/impacetevaluation/index.php>>.
- Montgomery, Tommie Sue. 1998. "International Missions, Observing Elections, and the Democratic Transition in El Salvador." In *Electoral Observation and Democratic Transitions in Latin America*, edited by Kevin J. Middlebrook, 53-90.
- National Research Council, Committee on Evaluation of USAID Democracy Assistance Programs. 2008. *Improving Democracy Assistance: Building Knowledge Through Evaluations and Research*. Washington, DC: The National Academies Press.
- Newman, Edward, and Roland Rich, eds. 2004. *The UN Role in Promoting Democracy: Between Ideals and Reality*. Tokyo: United Nations University Press.
- Nickerson, David W. 2005. "Scalable Protocols Offer Efficient Design for Field Experiments." *Political Analysis* 13(3): 233-252.
- Nielsen, Richard, and Daniel Nielson. 2008. "Lending Democracy: How Governance Aid May Affect Freedom." *Paper presented at the annual meeting of the APSA 2008 Annual Meeting*, Hynes Convention Center, Boston, Massachusetts.
- Nwankwo, Clement. 1999. "Monitoring Nigeria's Elections." *Journal of Democracy* 10 (4): 156-165.

- Oquaye, Mike. 1995. "The Ghanaian Elections of 1992 - A Dissenting View." *African Affairs* 94 (375): 259-275.
- Orozco, Manuel. *International norms and mobilization of democracy*. Ashgate, 2002.
- OSCE/ODIHR. Election Observation. 2005. *A decade of monitoring elections: the people and the practice*. Warsaw: OSCE Office for Democratic Institutions and Human Rights.
<http://www.osce.org/publications/odihr/2005/11/17148_478_en.pdf>.
- Ottaway, Marina. 1998. "Angola's Failed Elections." In *Postconflict Elections, Democratization, and International Assistance*, edited by Krishna Kumar. Lynne Rienner Publishers.
- Pastor, Robert A. 1998. "Mediating Elections" *Journal of Democracy* 9 (1): 154-163.
- Pevehouse, Jon C. 2005. *Democracy from Above: Regional Organizations and Democratization*. Cambridge: Cambridge University Press.
- Pevehouse, Jon C. 2002. "With a Little Help from My Friends? Regional Organizations and the Consolidation of Democracy." *American Journal of Political Science* 46 (3): 611-626.
- Qodari, Muhammad. 2005. "Indonesia's Quest for Accountable Governance." *Journal of Democracy* 16 (2): 73-87.
- Rich, Roland. 2001. "Bringing Democracy into International Law." *Journal of Democracy* 12 (3): 20-34.
- Robinson, William I. 1996. *Promoting Polyarchy: Globalization, US Intervention, and Hegemony*. Cambridge University Press.
- Schedler, Andreas, ed. 2006. *Electoral Authoritarianism: The Dynamics of Unfree Competition*. Lynne Rienner Publishers.
- Schedler, Andreas. 2002. "The Nested Game of Democratization by Elections." *International Political Science Review* 23(1): 103-122.
- Scranton, Margaret E. 1998. "Electoral Observation and Panama's Democratic Transition." In *Electoral Observation and Democratic Transitions in Latin America*, edited by Kevin J. Middlebrook, 53-90.
- Smith, Tony. 1994. *America's Mission: The United States and the Worldwide Struggle for Democracy in the Twentieth Century*. Princeton: Princeton University Press.
- Soremekun, Kayode. 1999. "Disguised Tourism and the Electoral Process in Africa: A Study of International Observers and the 1998 Local Government Elections in Nigeria." *Issue: A Journal of Opinion* 27 (1): 26-28.
- Wanandi, Jusuf. 2004. "The Indonesian General Elections, 2004." *Asia-Pacific Review* 11 (2): 115-131.
- Whitehead, Laurence, ed. *The International Dimensions of Democratization: Europe and the Americas*. Oxford: Oxford University Press, 1996.
- Wright, Joseph. 2009. "How Foreign Aid Can Foster Democratization in Authoritarian Regimes." *American Journal of Political Science* 53 (3): 552-571.