

# Improving Electoral Integrity with Information and Communications Technology

Michael Callen,<sup>\*</sup> Clark C. Gibson,<sup>†</sup> Danielle F. Jung<sup>‡</sup> and James D. Long<sup>§</sup>

## Abstract

Irregularities plague elections in developing democracies. The international community spends hundreds of millions of dollars on election observation, with little robust evidence that it consistently improves electoral integrity. We conducted a randomized control trial to measure the effect of an intervention to detect and deter electoral irregularities employing a nationwide sample of polling stations in Uganda using scalable information and communications technology (ICT). In treatment stations, researchers delivered letters to polling officials stating that tallies would be photographed using smartphones and compared against official results. Compared to stations with no letters, the letters increased the frequency of posted tallies by polling center managers in compliance with the law; decreased the number of sequential digits found on tallies – a fraud indicator; and decreased the vote share for the incumbent president in some specifications. Our results demonstrate that a cost-effective citizen and ICT intervention can improve electoral integrity in emerging democracies.

**Keywords:** Elections, electoral irregularities and fraud, randomized controlled trial, voting, information and communications technology, Uganda.

The study was approved by the UCSD IRB (Project #110178XX). The authors gratefully acknowledge the generous funding provided by the National Science Foundation (award #1118589). We thank Eli Berman, Eric Bjornlund, Sarah Bush, Karen Ferree, Tom Flores, Yusaku Horiuchi, Susan Hyde, Michael Kremer, Aila Matanock, Craig McIntosh, Pippa Norris, Maura O’Neill, Simon Osborn, Elinor Ostrom, Lauren Prather, Jake Shapiro, Alberto Simper, Charles Sprenger, and seminar participants at Harvard University, University of Washington, Indiana University, and the anonymous reviewers for their insightful comments. Nathan Combes, Maya Duru, Sara Kerosky, Doreen Lwanga, Laura Maxwell, Nicholas Obradovich, Amanda Pinkston, Arman Rezaee, Inbok Rhee, and Drew Wagstaff provided excellent research assistance. Long acknowledges support from the Harvard Academy for International and Area Studies. Callen and Jung acknowledge support by the AFOSR under Award No. FA9550-09-1-0314. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of AFOSR. The responsibility for any mistakes made in the manuscript lies with the authors.

<sup>\*</sup>Harvard University, Cambridge, MA 02138, USA, e-mail: [Michael\\_Callen@hks.harvard.edu](mailto:Michael_Callen@hks.harvard.edu)

<sup>†</sup>University of California at San Diego, La Jolla, CA 92093, USA, e-mail: [ccgibson@ucsd.edu](mailto:ccgibson@ucsd.edu)

<sup>‡</sup>Emory University, Atlanta, GA 30322, USA, e-mail: [danielle.jung@emory.edu](mailto:danielle.jung@emory.edu)

<sup>§</sup>University of Washington, Seattle, WA, 98105, USA and Harvard Academy for International and Area Studies, Cambridge, MA 02138, USA, e-mail: [jdlong@uw.edu](mailto:jdlong@uw.edu)

## INTRODUCTION

Free and fair elections, the cornerstone of modern democratic government, require officials to count and aggregate ballots accurately (Enikolopov et al. 2013; Klimek et al. 2012; Lehoucq 2003). The international community spends hundreds of millions of dollars every year trying to ensure fair elections in developing countries with widespread electoral malpractice (Bjornlund 2004). One common donor policy deploys election observers to improve electoral integrity (Bjornlund et al. 1992; Hyde 2007, 2010, 2011; Kelley 2012). Such missions, however, have not typically followed scientific designs to detect, nor have they had any consistent effects on, electoral malfeasance (Kelley 2012; Simpser and Donno 2012). This means that the widely publicized observer verdicts of whether an election is “free or fair” are based largely on anecdotal and unsystematic data.

While some research demonstrates positive effects from monitoring in certain settings (Enikolopov et al. 2013; Hyde 2007, 2010, 2011), other studies using cross-national data show that international efforts are actually more likely to be associated with null or negative impacts on electoral processes (Kelley 2012; Simpser and Donno 2012). This divergence may result from the protocols of standard observation missions, which typically do not measure their effect on electoral irregularities. First, observers are either sent to a non-random set of centers or deviate from randomized assignment protocols (by mistake or because they go to centers they believe more likely to experience fraud) (Enikolopov et al. 2013; Hyde 2007, 2010).<sup>1</sup> Election officials may predict observer deployments *ex ante* and strategically choose locations to commit fraud elsewhere. Second, observers pursue a variety of tasks during their missions, making it difficult to identify or measure any precise intervention. These diverse activities may include interacting with polling staff, voters, other observers, and party agents; inspecting voting materials; and watching voting and counting processes. Moreover, observers spend varying amounts of time on these activities. Third, observer missions typically do not employ precise outcome measures.<sup>2</sup> Since electoral irregularities, including fraud, come in a variety of forms, verdicts of “free and fair” implicitly refer to all of them while not necessarily measuring any of them systematically. Fourth, observation efforts rarely compare visited with non-visited polling centers, preventing an assessment of a monitoring treatment against an unmonitored control.<sup>3</sup> Finally, international observers seldom harness innovations in ICT. ICT allows users to increase government accountability and transparency by

<sup>1</sup>Hyde was the first scholar that we know of that sought to study the effects of randomized election observers. Enikolopov et al. (2013) constructed a sample by deploying domestic observers to 156 of 3,164 polling stations drawn randomly in Moscow.

<sup>2</sup>Hyde (2010) and Enikolopov et al. (2013) are exceptions. Hyde tests the effect of observer visits to polling stations by using unofficial government returns at the polling station level. Enikolopov et al. use official counts at the polling station level. Some domestic observer groups occasionally conduct parallel vote tabulations, which seek to employ random reports of polling station results to prevent fraud in the vote count.

<sup>3</sup>Hyde (2010) and Enikolopov et al. (2013) are exceptions: each compares vote counts at stations treated with observers to those not treated. In a different research design looking at the effect of observers on

quickly and cheaply aggregating diffuse information that is useful for auditing, such as polling station results, with the ability to scale coverage significantly beyond where international observers typically visit. Recent research also demonstrates how ICT can catalyze democratic mobilization through effective organization building and crowd-sourcing (Bond et al. 2012; Nisbet et al. 2012; Pickard et al. 2011; Valenzuela 2013). The shortcomings of standard election monitoring procedures thus hinder researchers and policymakers from accurately measuring the effect of observation on electoral integrity in a consistent, reliable, comparable, or scalable manner.

To address these challenges, we present results from a nation-wide randomized controlled trial (RCT) using a precise intervention and ICT designed to reduce irregularities and fraud associated with vote tallies at the polling center level conducted during Uganda's 2011 national election. Many fragile democracies like Uganda suffer from institutional weaknesses that foster bureaucratic corruption (Cohen 2008). Given the high return on rents from gaining office, unscrupulous political agents frequently threaten the integrity of electoral processes. They may do so locally by influencing the actions of polling center managers, or more centrally by using their control over the government's electoral machinery (Callen and Long 2015). Yet detecting and deterring such malfeasance is difficult since the illicit actions intended to manufacture electoral victories are often hidden.

Our intervention compares smartphone photographs of Declaration of Results forms—which are posted publicly at polling stations and list vote totals for each candidate—against the corresponding records published by the Election Commission at the conclusion of the election. In theory, this approach should allow us to perfectly observe any votes that are altered during the aggregation process. Our experiment involves delivering a randomized announcement of election monitoring, indicating the use of this smartphone technology. We measure the impacts of communicating information related to both monitoring and to punishments for malfeasance on election fraud. Specifically, we argue that in line with standard corruption models, the delivery of a letter announcing the monitoring of vote tallies by domestic researchers to polling center officials should increase the likelihood that ballots are correctly aggregated. We call this the “Monitoring” letter. This intervention derives from insights from the political economy of corruption that officials will change their expectation about the probability that potentially illegal behavior will be detected once monitoring is announced (Becker 1968; Becker and Stigler 1974; Björkman and Svensson 2009; Olken 2007; Olken and Pande 2013). Additionally, we investigate another treatment that reminds polling officials about the penalties for tally malfeasance, which we call the “Punishment” letter (Shleifer and Vishny 1993). Assuming some non-zero baseline probability that fraud is detected, announcing penalties could make them salient to election officials and therefore reduce the likelihood of irregularities. Including these two treatment arms allows us to compare the “Punishment” treatment to the “Monitoring” treatment.

the voter registration process, Ichino and Schündeln (2012) compare results from registration centers in treated and untreated electoral areas in four of Ghana's ten regions.

Moreover, since both messages in one letter may have an overall stronger effect, we also estimate the interaction of “Monitoring” and “Punishment.”

From our approach, we hypothesize

1. The incidence of electoral irregularities decreases in stations that received a “Monitoring” letter compared to control stations with no announcement (no letter delivery).
2. The incidence of election irregularities decreases in stations that received a “Punishment” letter compared to control stations with no announcement (no letter delivery).
3. The combination of “Monitoring” and “Punishment” messages decreases electoral irregularities to a greater degree than either message on its own.

Given that the incumbent president in a race has more institutional control over the electoral process than his or her rivals, we examine a fourth hypothesis:

4. The votes for the incumbent president decrease in stations that received any letter when compared to control stations.

## SAMPLE

Our sample included 1,095 total polling streams clustered within 1,001 total polling centers drawn from Uganda’s official national list of polling centers, by registered voters. Occasionally, polling centers will have more than one stream (for example, a school may be a polling center, with multiple voting locations, or streams, like individual classrooms). We consider any stream within a center to be “treated” since a polling center manager is responsible for all streams within a center. We sampled polling centers for this project using two separate protocols. In the Central, West, and East regions, we sampled polling centers using a multi-stage cluster design. In the first stage, we selected 25 counties, stratified by presidential margin of victory in the previous 2006 election. The two main competitors for office in 2011, incumbent president Yoweri Museveni and opposition leader Kizza Besigye, had both competed in 2006.<sup>4</sup> Specifically, we split counties into quintiles based on Museveni’s margin of victory and selected two counties in the most competitive quintile, three counties in the second quintile, five counties in the third quintile, seven counties in the fourth quintile, and eight counties in the least competitive quintile. We define “competitive” as the proportion of the vote that went to Museveni as reported by the Electoral Commission (EC). We predicted that fraud would more likely occur in areas that were less competitive given the ability of the incumbent

<sup>4</sup>President Yoweri Museveni of the National Resistance Movement (NRM) has governed Uganda since 1986, winning elections in 1996, 2001, 2006, and 2011. He won in 2011 with 68 percent of the vote, and his NRM took 70 percent of seats in parliament. The second place candidate, Kizza Besigye of the Forum for Democratic Change (FDC), garnered 26 percent of the vote.

Museveni to exercise greater political control over electoral institutions and polling center officials in his areas of strong support.<sup>5</sup> Therefore, our approach oversampled less competitive counties. In the second stage, we drew 95 subcounties in total from within the 25 sampled counties. The number of subcounties per county ranges from 2 to 15, proportionate to the number of registered voters. In the third and final stage, we drew 601 total polling centers with about 5.8 polling centers in each subcounty. Our sample for the Western, Central, and Eastern regions therefore consisted of 25 counties, 95 subcounties, and 601 polling centers.

In the North, we sampled only polling centers in subcounties that experienced violence during the war between the Ugandan Government and the Lord's Resistance Army (LRA) over the period 1987–2008 based on data from Armed Conflict Location and Event Dataset (ACLED). We sampled the North differently than the rest of the country because our funder required that we focus our study only in subcounties affected by conflict, so non-conflict Northern subcounties had zero probability of selection.<sup>6</sup> This requirement was later removed. We therefore also present results from the North separately from the Central, West, and East regions to analyze our data across the entire national sample, while maintaining the representativeness of the sample that excludes the Northern region. 30 counties, 56 subcounties, and 400 polling centers comprise our Northern sample.<sup>7</sup>

## TREATMENT

Our interventions consisted of letters delivered by researchers to polling center officials during voting on election day. Three versions of the letter's message constituted the treatment arms of the experiment: (1) The "Monitoring" letter informed officials that researchers would return to take a photograph of the Declaration of Results form—the tally—that electoral law requires managers to post publicly at the polling center after ballot counting, and that these results would be compared with the center's final count as certified by the Ugandan EC. (2) The "Punishment" letter reminded polling center managers of the 2.4 million Ugandan

<sup>5</sup>We note that we could not rely upon data directly measuring irregularities from the previous 2006 election to stratify our sample. First, lacking an intervention such as ours, we are unaware of any researchers who directly measured irregularities at the polling station level using our outcome variables. Second, electoral units and boundaries changed significantly between both elections at the subcounty and polling station level, making any comparisons impossible. Hence, using margin of victory from 2006 at the county level was the best alternative.

<sup>6</sup>Specifically, the sampling frame in the North comprised subcounties which had either (i) recorded a violent incident in the ACLED dataset or (ii) which shared a border with a subcounty that records violence in the ACLED dataset. Subcounties not satisfying either (i) or (ii) had no probability of selection. Polling centers were randomly selected from the universe of polling centers in selected subcounties. This procedure precluded the development of sampling weights in the Northern subsample.

<sup>7</sup>There are 41 possible counties and 228 possible subcounties in the North. The sampling protocol in the North does not allow us to construct weights as described in Footnote 6. We indicate in each table when the North sample is either included or excluded from the analysis sample. Weighted results for Central, West, and East are available upon request.

shilling fine (approximately 1,000 USD, not purchasing power parity adjusted) and/or five years imprisonment for inaccurately reporting voting returns. (3) The third letter included the text of both the monitoring and punishment treatments. Control centers received no letter. We include the text of the “Monitoring” message, the “Punishment” message, and “Both” messages versions of the letters in Appendix A, Figures A1–A3. We randomized the ordering of the “Monitoring” and “Punishment” messages in the “Both” messages version to test for possible order effects. We divided the total sample to include 227 centers in each treatment arm, and 320 centers in the control, based on power calculations of anticipated treatment effects. We sent researchers to photograph tallies of all centers in the sample the day after the election.

To assign treatment status, we conducted a baseline survey in late January 2011, approximately one month before the election. The survey used the polling centers in the experimental sample as primary sampling units (PSUs), so that the survey sampling followed the experimental sampling described above. Enumerators instituted random walk patterns from the polling center location, selecting every fourth house on the left. Within households, enumerators randomly selected voting eligible (18 years or older) respondents via Kish grid. The survey has a 56–44 percent male to female ratio. Enumerators interviewed five respondents per polling center catchment area, yielding a total national sample of 5,007 respondents. The survey probed questions on political attitudes and perceptions, as well as demographic covariates, as shown in Table A1 in the Appendix. Table A1 also checks the balance of these measures across treatment arms.

We stratified our treatment assignment on (i) county, (ii) the share of survey respondents on our baseline survey who indicated that they had read a newspaper in the last seven days, and (iii) on the share who indicated that Museveni’s performance as president had been “good” or “excellent.” After randomly assigning treatment status, we assessed covariate balance between treatment and control polling centers. In instances where differences in covariate means were statistically significant below a threshold of  $p = 0.25$ , we instituted a procedure of re-randomization with 691 iterations. This so-called “big stick” approach follows the re-randomization procedures described by Bruhn and McKenzie (Bruhn and McKenzie, 2009).<sup>8</sup>

We provided letters, smartphones, and training to 370 Ugandan researchers and field managers. After being credentialed by the EC as an election observer group as the University of California, each member of the group applied for—and received—accreditation as official observers, certifying them to carry out the research although the EC did not know the substance of the study or the sample of polling centers. We furnished each researcher with an HTC Android-compatible

<sup>8</sup>We balanced on all of the covariates in Table A1. We randomized until the minimum  $p$ -value was 0.25. This “big stick” approach is standard in field randomized control trials. Bruhn and McKenzie provide a comprehensive treatment, based on simulations, of how this affects the statistical power of randomized control trials. They conclude that “as ye randomize, so shall ye analyze,” essentially arguing that one must include block fixed effects. Our results are robust to both including and excluding the effects.

smartphone, equipped with a custom application designed by Qualcomm, Inc. that allowed the photo capture of tallies and the completion of a short survey. Appendix Figure A4 displays the application. Researchers delivered the treatment letters to polling center managers on February 18 (election day) and returned to photograph tallies on February 19 when they were legally required to be posted publicly. We note near universal compliance with letter delivery save for one polling center that fell out of sample on election day, with 95 percent of polling center managers acknowledging receipt by agreeing to sign a copy of the delivered letter. Each tally photo had a unique polling center identification number, as well as a time and date stamp embedded in the data packet to ensure compliance. Figure A5 shows an example of an actual tally photo. After entry on the smartphone, data were immediately sent via Uganda's cellular networks to a server located at the University of California, San Diego. If the smartphones did not carry a signal due to poor network coverage, the application sent the data packet once within network range.

## RESULTS

We measure the effect of the intervention by comparing vote totals from treated and control polling centers on two sets of data: the official count at the polling center level produced by the EC and differences in votes in our set of photographed tallies compared to EC results. Given that potentially corrupt polling managers will try to hide illicit behavior, we employ three measures to determine irregular electoral activities that we can observe, which can range from unknowingly committed illegal administrative acts to intentional fraud.<sup>9</sup> First, we use instances of the absence of provisional tallies, which we label as “Missing.” The handbook that the EC used to train presiding officers clearly states that Ugandan law requires the posting of tallies, and that failure to do so is an abdication of duty and can result in official censure.<sup>10</sup> The election law requires tallies to be posted publicly to prevent efforts to rig the election through an obviously manipulated local ballot count or in the process of changing the count in the aggregation process.<sup>11</sup> Data for “Missing” come from our researchers' visits to polling centers to document the presence or absence of tallies. 77.5 percent of our sample had missing tallies, demonstrating widespread violations of Uganda's election law, which left most Ugandans unable to know or verify the count at their own polling center. Second, we record whether the last two digits in the winning candidate's vote total are adjacent, which we label “Adjacent.” The analysis of adjacent digits—a commonly used measure of electoral fraud—comes from studies in behavioral psychology that document the tendency of humans to disproportionately use adjacent digits when manufacturing numbers (Beber and Scacco 2012). If election officials are inventing numbers when writing down vote

<sup>9</sup>For more details about the justification of these outcome measures, please see the extended discussion in Callen and Long (2015).

<sup>10</sup>See Presidential Elections (Amendment) Act (2010) Section 51 subsection 2.

<sup>11</sup>For more details about the relationship between the failure to post, or removal of, tallies and fraud, see Callen and Long (2015).

*Table 1*  
**OLS Estimates of the Effects of the Treatment Letters on Missing Tallies, Adjacent Digits, and Votes for Museveni**

	(1)	(2)	(3)	(4)	(5)	(6)
	Missing	Adjacent	Missing	Adjacent	Museveni votes (ln)	Museveni votes (ln)
<b>Monitoring</b>	− 0.106** (0.046) [0.020]	− 0.080* (0.047) [0.076]	− 0.055 (0.033) [0.040]	− 0.068** (0.034) [0.036]	− 0.025 (0.027) [0.164]	− 0.031 (0.023) [0.088]
<b>Punishment</b>	− 0.101** (0.048) [0.012]	− 0.058 (0.050) [0.100]	− 0.055 (0.035) [0.056]	− 0.103*** (0.033) [0.004]	0.006 (0.024) [0.620]	− 0.005 (0.021) [0.392]
<b>Both</b>	− 0.117*** (0.041) [0.012]	− 0.074 (0.045) [0.080]	− 0.060* (0.033) [0.028]	− 0.065* (0.034) [0.052]	0.013 (0.023) [0.660]	0.015 (0.021) [0.764]
<b>Total votes (ln)<sup>+</sup></b>	0.000 (0.037)	− 0.012 (0.046)	0.012 (0.026)	− 0.005 (0.035)	1.026*** (0.026)	0.887*** (0.036)
<b>Constant</b>	0.779 (0.213)	0.309 (0.267)	0.721 (0.148)	0.254 (0.199)	− 0.722*** (0.151)	0.099 (0.151)
<b>North included</b>	No	No	Yes	Yes	Yes	Yes
<b>Outliers excluded</b>	No	No	No	No	No	Yes
<b>Controls included</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b>Observations</b>	556	556	1,028	1,025	1,026	926
<b>R-squared</b>	0.492	0.180	0.395	0.159	0.885	0.869
<b>Clusters<sup>++</sup></b>	517	517	913	911	911	828
<b>Joint significance<sup>+++</sup></b>	0.001	0.060	0.030	0.004	0.623	0.384
<b>Any letter</b>	− 0.108*** (0.033) [0.016]	− 0.071* (0.037) [0.028]	− 0.056** (0.026) [0.004]	− 0.079*** (0.027) [0.004]	− 0.002 (0.018) [0.048]	− 0.007 (0.016) [0.040]
<b>Constant</b>	0.778 (0.213)	0.308 (0.266)	0.721 (0.148)	0.245 (0.199)	− 0.720*** (0.149)	0.094 (0.153)
<b>R-squared</b>	0.492	0.179	0.395	0.157	0.885	0.868

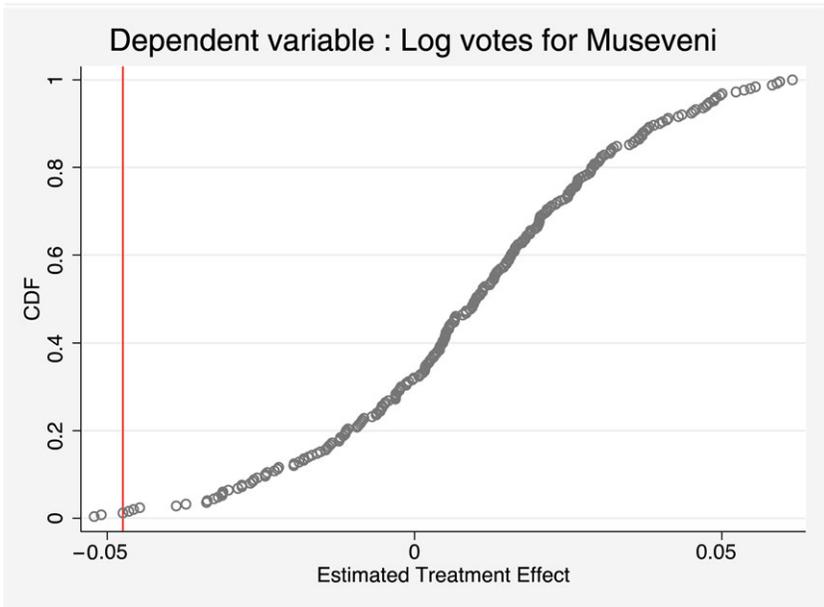
*Notes:* Estimation is conducted with stratum fixed effects. Coefficients in the top panel are the point estimates for effect of each type of letter intervention. <sup>+</sup>Log of total votes at each station accounts for station size. <sup>++</sup>Standard errors are clustered at the polling center level to account for the intra-class correlation of errors within polling centers (where multiple polling stations vote at a single location, typically a school or church). Fisher Exact Test *P*-values corresponding to the Sharp Null of no treatment effect for any unit are reported in brackets. Fisher Exact Tests are calculated based on 250 hypothetical treatment assignments following the same assignment protocol as the actual assignment. “Missing” is a dummy variable equal to one if the Declaration of Results form (the tally) was missing at the polling center when our interviewers visited to take a photograph. “Adjacent” is a dummy variable equal to one if the last two digits in the vote total for the top vote recipient at a given polling center are adjacent (e.g. 34 or 21) in the EC’s official report of votes. *P*-values corresponding to a standard difference in means test are denoted by superscript stars: \*\*\**p* < 0.01, \*\**p* < 0.05, \**p* < 0.1. For example, in column 1, the “Monitoring” letter reduced the number of missing tallies by 10.6 percentage points at the *p* < 0.05 level. <sup>+++</sup>Test of joint significance: probability that the null hypothesis can be rejected. Sample sizes differ by polling center and stream due to data availability of tallies and EC certified results. The lower panel reports estimate for “Any letter” being delivered. Note that the specification of the model otherwise is identical, as are the number of clusters and observations.

totals rather than recording the actual ballot count, they will likely use numbers with adjacent digits more frequently than at random. The data for the “Adjacent” tests come from the EC’s official election results. We note that adjacent digits produced at random should occur 18 percent of the time, but in our sampled polling centers, occurred much more frequently (see Table 1). Third, following other studies that examine the effects of monitoring on votes for powerful candidates (Callen and Long 2015; Hyde 2007, 2010), we use the log of the total votes for the incumbent

president Yoweri Museveni (“Museveni Votes”), the candidate most likely to benefit from rigging given his executive control over electoral administration. Descriptive statistics for these outcome variables can be found in Appendix Table A4. We estimate the treatment effects on our three dependent variables using six separate linear regressions, with standard errors clustered by polling center and stratum fixed effects. Intercepts represent the mean outcome in the control sample, with coefficients showing the percentage point change in the likelihood of the outcome resulting from treatment.

Our tests indicate that the letter interventions reduced irregular electoral activities. The results found in column 1 of [Table 1](#) (corresponding to Appendix Figure A6, panel A), shows that the “Monitoring” letter decreased missing tallies by 10.6 percentage points in the non-North sample. [Table 1](#) column 2, which uses adjacent digits as the fraud indicator on the same sample, shows that point estimates are always negative although less significant: “Monitoring” decreased adjacent digits by 8 percentage points, “Punishment” by 5.8 percentage points, and “Both” by 7.4 percentage points (corresponding to Appendix Figure A6 panel B). Column 3 in [Table 1](#) uses the entire sample and shows a significant reduction of 6 percentage points of missing tallies for the “Both” letter, and insignificant but negative point estimates for “Monitoring” (5.5 percent) and “Punishment” (5.5 percent) (corresponding to Appendix Figure A6 panel C). [Table 1](#) column 4 displays the effect of the letters over the entire sample on the presence of adjacent digits. All three letters reduced adjacency at a significant level: “Monitoring” by 6.8 percentage points, “Punishment” by 10.3 percentage points, and “Both” by 6.5 percentage points (corresponding to Appendix Figure A6 panel D). There are no consistent patterns between the effects of the three letters. Thus, while all three treatment letters had an effect, the results do not indicate that one treatment is unambiguously stronger than another’s, or that they interact. We therefore pool our treatments, testing the impact of “Any Letter” in the lower panel of [Table 1](#). Delivering any letter reduces missing tallies by 10.8 percentage points in the non-North and 5.6 percentage points in the entire sample (specifications 1 and 3 in the lower panel of [Table 1](#)). Similarly, adjacent digits are reduced by 7.1 percentage points in the non-North and 7.9 percentage points in the entire sample (specifications 2 and 4 in the lower panel of [Table 1](#)).

We report the effects of the treatment on the log of Museveni votes in columns 5 and 6 in [Table 1](#). Column 5 shows no effect from the three treatments at standard levels of significance. This null finding for hypothesis 4 may arise from the ineffectiveness of monitoring in polling centers under the control of the incumbent Museveni or of one of his competitors. When we exclude the polling centers in the bottom and top five percentiles of the log of Museveni votes from the sample, the “Monitoring” letter decreased Museveni’s vote share by more than 3 percentage points (column 6), and we reject the Sharp Null of no effect with a  $p$ -value of 0.088, providing some weak support for hypothesis 4. We note that the small effect of the “Monitoring” treatment on Museveni’s votes may arise from the lack of a credible threat posed by reminders of penalties to polling officials who may alter votes on his behalf.



*Figure 1*  
(Color online)

*Notes:* This figure plots 250 estimated treatment effects estimated by regressing the log of votes cast for Museveni at a given polling center on a dummy equal to one for polling centers assigned to treatment hypothetically using a randomization protocol identical (except for the random number seed) to the actual protocol used to assign treatment and on the log of votes cast at the polling center. The vertical line reflects the estimated treatment effect of  $-0.047$ , which yields a Fisher exact  $p$ -value of 0.016. This same approach is taken to estimate the Fisher exact  $p$ -values reported in Table 1.

Beneath the estimated coefficients corresponding to the treatment effect of “Monitoring”, “Punishment,” and “Both” in Table 1 we also report  $p$ -values corresponding to the Fisher Exact Test of the Null of no treatment effect for any unit. The randomization inference approach we use, therefore, tests a hypothesis that is distinct from the standard null of no average effect corresponding to a difference in means test. The Fisher  $p$ -value is calculated by comparing the treatment effect estimated based on the actual assignment with treatment effects given by hypothetical assignments using the same randomization protocol. Figure 1 demonstrates this approach by plotting the position of our estimated treatment effect of  $-0.047$  log points as a vertical line in the cumulative distribution of simulated treatment effect estimates. Of 250 simulated treatment assignments, two produce estimates of a more negative treatment effect, yielding a Fisher  $p$ -value of 0.016. We take this approach as it addresses two challenges for inference in our setting. First, because the  $p$ -value is calculated based on a comparison of hypothetical assignments obtained using the same randomization protocol, we can be confident that our result is not an artifact of our specific assignment protocol. Second, because the approach is non-parametric and the  $p$ -value simply reflects where the actual estimate sits in the distribution of potential estimates, we can perform inference without making assumptions on the appropriate form of the standard errors, including the level of

clustering. We reject the Sharp Null in every case where we reject the null from a difference in means test. We also reject the Sharp Null for six additional estimates in the top panel of [Table 1](#) where we do not reject the null of no average treatment effect at the 10 percent level.

Figure A7 shows the effects of the letters on Museveni's vote total, using only the sample for which we have photographs comparing those results to the EC data. Recall that a majority of tallies were not posted in our treatment sample, generating the smaller and non-random sample we use in Figure A7. Given treatment-related attrition in this sample, we cannot estimate treatment effects across the entire sample using the results from the tests found in [Table A3](#) in Appendix A.<sup>12</sup> These bounds include zero, but only under the extreme assumption that treatment induced presiding officers to post tallies in polling centers where Museveni received the largest number of votes in the observed treatment sample. There is considerable fraud in this sample. In Figure A7, the lighter bars show the difference between the treatment and the control groups on photographed tallies in this non-representative sample; the darker bars represent the difference on the official results published by the EC. The point estimates are consistent with treatment reducing the number of votes Museveni received as photographed in the tallies. In the sample excluding the North, the "Monitoring" letter reduced Museveni votes by about 30 from an average of 307 votes per polling center; the "Punishment" letter by 26 votes, and "Both" by almost 49 votes—a decrease of 16 percent.

The darker bars in Figure A7 show that the reduction in Museveni's vote due to the treatment letters is still detectable after the EC received and amended the tallies to produce their official, certified results. Under all three treatments, however, the decrease in votes for the incumbent is less than the effect found in photographed tallies. This result could reflect attempts by officials in higher echelons of the electoral administration to offset reductions caused by the treatment letters to Museveni's vote totals after receiving the polling station counts, although we have no additional evidence to support this possibility.<sup>13</sup>

## DISCUSSION

We believe our findings provide insights for theory and policy, with application to elections in other emerging democracies. We contribute to studies regarding

<sup>12</sup>Given the non-random nature of the photo sample, we conducted a robustness check using the Lee Bounding technique to estimate the effect of treatment in the presence of non-random attrition. [Table A2](#) in the Appendix provides estimates using this method using the absolute value of the difference between the number of votes recorded for Museveni on the tally and the number recorded in the administrative record.

<sup>13</sup>We note a lack of geo-coded locations for polling centers in Uganda, making it impossible to assess fully the presence and extent of spatial spillovers due to our treatments. However, as noted in the results above, we see evidence of a vote recovery strategy due to our intervention at work in the Appendix, [Table A3](#).

efforts to reduce corruption (Becker 1968; Becker and Stigler 1974; Björkman and Svensson 2009; Ferraz and Finan 2008; Olken 2007; Olken and Pande 2013; Shleifer and Vishny 1993), and improve electoral integrity. We provide evidence that the intervention letters: (1) decreased the illegal practice of not publicly posting tallies; (2) reduced the number of adjacent digits found on tally sheets; and (3) reduced the vote share for the incumbent President Museveni under certain specifications. The data also show the letter that included both monitoring and punishment messages did not consistently strengthen the treatment effect, and that the decreases in polling center votes for Museveni caused by the letters is systematically related to the inclusion of additional votes at the level of the EC.

Our study also provides important lessons for future election observation efforts. Despite their considerable cost, standard international election missions lack consistent scientific evidence to claim that they improve electoral integrity. While such missions may provide other benefits, such as offering international support for a democratic process, election observation per se does not typically employ research designs that precisely measure their impact on irregular or illicit electoral activities in a representative sample of polling locations. Our study presents a novel, cost-effective, and scalable technique to do so. It contributes to existing approaches by combining experimental design and ICT to enhance the efforts of groups seeking to detect and reduce illicit behavior.

We conclude with some general observations on the potential cost-effectiveness of this approach relative to standard missions. While those missions have a broader set of capacity building objectives beyond reducing election fraud, we believe these comparisons remain salutary. The European Union (EU) provides the total number of polling stations and expenditure data for only two of its 142 mission reports (Guinea-Bissau, 2009 and Nigeria, 2009), with each costing over \$20,000 per polling station, although these high costs could include other types of electoral support (European Union 2013). The EU reports that it spends approximately \$4 million on an average observation mission (European Union 2006). While the EU declined to publicize their budget for their mission to the 2011 election in Uganda, they report visits to 643 polling centers and tally centers (European Union 2011). Using their average budget, we calculate a \$6,220 cost per polling center, with no scientific evidence of impact documented in their report. With a project budget of \$40,000 and 1,001 sampled polling centers, our price per polling center was about \$40. We note that our design can be adopted easily across a range of countries, and while it would likely be most suitable for domestic groups that can scale their efforts beyond the generally smaller international observation teams, it also points to design features that these groups could use to generate systematic information upon which to base their judgments about election quality. We note that we designed our intervention to target certain behaviors, and any increase in scaling could temporarily off-set gains made at observed polling stations by displacing fraud to unobserved polling stations or to other rigging strategies. For these reasons, we believe scaling with ICT via citizen-based viral adoption may prove the most fruitful avenue to obtain

near-universal monitoring coverage of polling centers in a country to report on a variety of illicit activities.

## SUPPLEMENTARY MATERIAL

For supplementary material for this article, please visit Cambridge Journals Online: <http://dx.doi.org/10.1017/XPS.2015.14>.

## REFERENCES

- Beber, B., and A. Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20(2): 211–34.
- Becker, G. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76: 169–217.
- Becker, G., and G. Stigler. 1974. "Law Enforcement, Malfeasance, and Compensation of Enforcers." *Journal of Legal Studies* 3(1): 1–18.
- Björkman, M., and J. Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124(2): 735–69.
- Bjornlund, E. 2004. *Beyond Free and Fair: Monitoring Elections and Building Democracy*. Washington, DC: Woodrow Wilson Center Press and Baltimore, MD: Johns Hopkins Univ. Press.
- Bjornlund, E., M. Bratton, and C. Gibson. 1992. "Observing Multiparty Elections in Africa: Lessons from Zambia." *African Affairs* 91(364): 405–31.
- Bond, R., C. Fariss, J. Jones, A. Kramer, C. Marlow, and J. Settle. 2012. "A 61-million-person Experiment in Social Influence and Political Mobilization." *Nature* 489(7415): 295–98.
- Bruhn, M., and D. McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1(4): 200–32.
- Callen, M., and J. D. Long. 2015. "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan." *American Economic Review* 105(1): 354–81.
- Cohen, J. 2008. "Uganda Confronts Corruption, Slowly." *Science* 321(5888): 522–25.
- Enikolopov, R., V. Korovkin, M. Petrova, K. Sonin, and A. Zakharov. 2013. "Field Experiment Estimate of Electoral Fraud in Russian Parliamentary Elections." *Proceedings of the National Academy of Sciences of the United States of America* 110(2): 448–52.
- European Union. 2006. "European Union: External Action FAQ." [www.eeas.europa.eu/](http://www.eeas.europa.eu/), accessed November 4, 2013.
- European Union. 2011. "Final Report on Uganda's General Elections 18 February 2011." ([http://eeas.europa.eu/eucom/pdf/missions/eucom\\_uganda2011\\_final\\_report\\_en.pdf](http://eeas.europa.eu/eucom/pdf/missions/eucom_uganda2011_final_report_en.pdf)), accessed November 4, 2013.
- European Union. 2013. "EU Election Observation Missions." (<http://www.eucom.eu/home>), accessed November 4, 2013.
- Ferraz, C., and F. Finan. 2008. "Exposing Corrupt Politicians: The Effect of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123(2): 703–45.
- Hyde, S. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 60(1): 37–63.

- Hyde, S. 2010. "Experimenting in Democracy Promotion: International Observers and the 2004 Presidential Elections in Indonesia." *Perspectives on Politics* 8(2): 511–27.
- Hyde, S. 2011. *The Pseudo-Democrat's Dilemma: Why Election Monitoring Became an International Norm*. Ithaca, NY: Cornell University Press.
- Ichino, N., and M. Schündeln. 2012. "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana." *Journal of Politics* 74(1): 292–307.
- Kelley, J. 2008. "Assessing the Complex Evolution of Norms: The Rise of International Election Monitoring." *International Organization* 62(2): 221–55.
- Kelley, J. 2012. *Monitoring Democracy: When International Election Observation Works and Why it Often Fails*. Princeton, NJ: Princeton Univ. Press.
- Klimek, P., Y. Yegorov, R. Hanel, and S. Thurner. 2012. "Statistical Detection of Systematic Election Irregularities." *Proceedings of the National Academy of the Sciences of the United States of America* 109(41): 16469–73.
- Lehoucq, F. 2003. "Electoral Fraud: Causes, Types, and Consequences." *Annual Review of Political Science* 6(1): 233–56.
- Nisbet, E., E. Stoycheff, and K. Pearce. 2012. "Internet use and Democratic Demands: A Multinational, Multilevel Model of Internet use and Citizen Attitudes about Democracy." *Journal of Communication* 62(2): 249–65.
- Olken, B. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200–49.
- Olken, B., and R. Pande. 2013. Governance Review Paper J-PAL Governance Initiative MIT. (<http://www.povertyactionlab.org/publication/j-pal-governance-initiative-review-paper>), accessed August 25, 2015.
- Pickard, G., I. Rahwan, W. Pan, M. Cebrian, R. Crane, A. Madan, and A. Pentland. 2011. "Time-critical Social Mobilization." *Science* 334(6055): 509–12.
- Shleifer, A., and R. Vishny, R. 1993. "Corruption." *Quarterly Journal of Economics* 108(3): 599–617.
- Simpser, A., and D. Donno. 2012. "Can International Election Monitoring harm Governance?" *Journal of Politics* 74(2): 501–13.
- Valenzuela, S. 2013. "Unpacking the use of Social Media for Protest behavior: The Roles of Information, Opinion Expression, and Activism." *American Behavioral Scientist* 57(7): 920–42.