Economics 215B
(Lecture 2)

Ethan Kaplan

February 8, 2008
1 Power Calculations

\[ Y_i = \alpha + \beta T_i + \epsilon_i \]

OLS Estimator is:

\[
\min_{\alpha, \beta} \sum_{i=1}^{I} (Y_i - \alpha - \beta T_i)^2
\]

\[
\alpha : -2 \sum_{i=1}^{I} (Y_i - \alpha - \beta T_i) = 0
\]

\[
\beta : -2 \sum_{i=1}^{I} (Y_i - \alpha - \beta T_i) T_i = 0
\]

This implies the following estimators for the treatment effect \( \beta \) :

\[
\hat{\beta} = \frac{\sum_{i=1}^{I} (Y_i - \bar{Y}) (T_i - \bar{T})}{\sum_{i=1}^{I} (T_i - \bar{T}) (T_i - \bar{T})}
\]
We can also compute the Standard Error by taking the $V(\hat{\beta})$: (and remembering that $P(T = 1) = P$)

$$SE(\hat{\beta}) = \sqrt{S^2 (X'X)^{-1}}$$

$$= \sqrt{\frac{\sigma_\varepsilon^2}{P(1-P)N}}$$

- **Size and Power**
  - Size of a Test: Probability of a Type I Error (Probability of Failing to Reject a True Null) = 1 - Confidence Level.
  - Power of a Test: 1 - Probability of a Type II Error
(Probability of Rejecting a False Null Hypothesis)

<table>
<thead>
<tr>
<th>Do Not Reject $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Decision</td>
<td>Type I Error</td>
</tr>
<tr>
<td>$H_0$ is True</td>
<td>Size of Test</td>
</tr>
<tr>
<td>$1 - \alpha$ : Confidence Level</td>
<td>$\alpha$ : Significance Level</td>
</tr>
<tr>
<td>$H_0$ is False</td>
<td>Type II Error</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Correct Decision</td>
</tr>
<tr>
<td>$1 - \omega$ : Power of Test</td>
<td></td>
</tr>
</tbody>
</table>

- In order to reject a null hypothesis of no effect at an $\alpha$ level of confidence, we need:

$$\hat{\beta} > t_{\alpha} SE \left( \hat{\beta} \right)$$

- If we want power of $1 - \omega$ :

$$\hat{\beta} > \left( t_{1-\omega} + t_{\alpha} \right) SE \left( \hat{\beta} \right)$$

$$SE \left( \hat{\beta} \right) = \sqrt{V \left( \hat{\beta} \right)} = \sqrt{\frac{\sigma^2}{P \left( 1 - P \right) N}}$$
• Therefore the Minimum Detectable Effect (where $\alpha$ is the size and $1 - \omega$ is the power):

$$\hat{\beta} > (t_{1-\omega} + t_{\alpha}) \sqrt{\frac{\sigma^2}{P (1 - P) N}}$$
• Treatment/Control Balance: Goal - maximize MDE choosing proportion of

$$\min_{N,P} \left( t_{1-\omega} + t_{\alpha} \right) \sqrt{\frac{\sigma^2_{\epsilon}}{P (1 - P) N}} - \lambda [N (1 - P) c_c + NP c_t - B]$$

  - Solution:

$$\frac{P}{1 - P} = \sqrt{\frac{c_c}{c_t}}$$

• Group Effects: Now suppose we have grouped data with group effects: $v_j$. Then, we estimate:

$$Y_{ij} = \alpha + \beta T_{ij} + v_j + \epsilon_{ij}$$

  - where there are $J$ clusters of size $n$

  - $v_j \sim i.i.d. N \left( 0, \tau^2 \right)$ and $\epsilon_{ij} \sim i.i.d. N \left( 0, \sigma^2_{\epsilon} \right)$

  - Then we get as our MDE:

$$\sqrt{\frac{1}{P (1 - P)}} \sqrt{\frac{n \tau^2 + \sigma^2_{\epsilon}}{nJ}}$$
With individual randomization, we would get:

\[
\sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\tau^2 + \sigma^2}{nJ}}
\]

Ratio between the two = \(D = \sqrt{1 + (n - 1) \rho}\)

where \(\rho = \frac{\tau^2}{\tau^2 + \sigma^2}\)

With imperfect compliance:

\[
\sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{Nc - s}} \frac{1}{\sqrt{\sigma^2}}
\]


\[V(\text{Uncond.}) > V(\text{Cond.}) > V(\text{Stratified})\]
– Generally (with spherical disturbances): Standard Errors Given By:

\[ \sigma^2_\varepsilon \left( X'X \right)^{-1} \]

– With stratification, this is a diagonal matrix in which case adding a dimension of stratification (constructed to be orthogonal to the other dimensions) will always reduce the standard errors.
• Stratification:

1. No effect on $\beta_T$ of inclusion of strata in regressions

$$\beta_T = \left[T'(I - P)T\right]^{-1} T'[I - P]y$$

where

$$P = X \left(X'X\right)^{-1} X'$$

but since

$$T'X = 0$$

we get:

$$T'(I - P) = T - 0 = T$$

and thus

$$\beta_T = \left[T'T\right]^{-1} T'y$$

2. Ambiguous Effect on Standard Error but in practice a reduction of standard error

(a) Lowers $\sigma^2_\varepsilon$
(b) Doesn’t raise $(X'X)^{-1}$ with full stratification; could in theory raise $(X'X)^{-1}$ in small samples without stratification.

(c) Could raise degrees of freedome correction.

3. Allows for greater number of subgroups and thus greater manipulability (randomly finding a subgroup with a significant effect but a true null of hypothesis of no effect).

4. Allows for looking at heterogeneity in treatment effect across different covariates.
2 Intention to Treat Estimates in Experiments

- Treatment: $T$, Assignment of Treatment: $Z$

- Average Treatment Effect (ATE):
  \[ E(Y_i^T - Y_i^C) \]
  - If you can actually randomize treatment

- Intention To Treat (ITT):
  \[ E(Y_i^T - Y_i^T | Z) = E(Y_i^T | Z = 1) - E(Y_i^T | Z = 0) \]
  - If you can randomize access to Treatment but not Treatment itself

- Do we want the intention to treat estimate or the treatment effect?
3 Gerber and Green: The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment

- Could ask people at voting polls if they were contacted (in person, by phone or via mail) but selection bias.

- Gerber and Green randomize access to contact (in person, phone or mail). Control is then given by:

\[ T_C = \alpha T_R + (1 - \alpha) T_{NR} \]
• Treatment is given by:

\[ T_T = \alpha (T_R + t) + (1 - \alpha) T_{NR} \]

• The estimate of \( t \) (the treatment effect) is then:

\[
\begin{align*}
T_t - T_C &= \alpha t \\
t &= \frac{T_t - T_C}{\alpha}
\end{align*}
\]

• Randomized contacts independently (mailing, phone, in person contact). Therefore, can look at interactions.

  – Benefit of independent randomization: Can look at interactions between types of contact.

  – Costs of independent randomization: Less power for each type of interaction.
• Implementation

  – Eliminate:

    * Students: why?

    * PO Boxes: why?

  – Mailing: 0, 1, 2, or 3 messages

  – Phone: Survey Company

  – Canvassing: Graduate Students
• What do they estimate:

\[ ITT : \]
\[ T_C - T_T \]

• IV Estimate:

\[ IV : \]
\[ \frac{T_C - T_T}{\alpha} \]

• Computation of \( \alpha \):

\[
\alpha = 25.3\% \text{ for phone} \\
= 28.\% \text{ for visit} \\
= 12.4\% \text{ for both} \\
= ? \text{ for mail?}
\]

• Alternative way of writing IV:

\[
T_i = \alpha + \beta C_i + \epsilon_i \\
C_i = \gamma Z_i + \delta_i
\]
where $T_i$ is turnout, $C_i$ is contact and $Z_i$ is intention to treat.

- why is there no constant term in the first stage regression?
findings indicate that personal canvassing is highly effective, much more so than the direct mail and telemarketing campaigns that have come to displace it. The implication is that the decline in voter turnout may be due to the changing character of American campaigns. Although the volume of mobilization activity remains considerable, its increasingly impersonal nature draws fewer people to the polls.

**EXPERIMENTAL DESIGN**

This field experiment was conducted in New Haven, Connecticut, which has a population of approximately 100,000. In September 1998 we obtained a complete list of registered voters, from which we created a data set of all households with one or two registered voters. To eliminate students from the sample, all names with post office box addresses were excluded, as was one voting ward that encompasses a university and student housing. We were left with 29,380 individuals (22,077 households), whose participation in the 1998 election could be determined from public records.

Our study was designed to measure the effect of personal canvassing, telephone calls, and direct mail appeals on voter turnout. Through a series of random assignments, the sample was divided into control and experimental groups. Table 2 shows the sample sizes of each group for the $2 \times 2 \times 4$ design.3 The treatment and control groups for the three experiments overlap, such that 10,800 people were assigned no intervention; 7,369 were sent at least one mailing but received no other appeal; 2,686 were slated only for personal contact; and 958 were assigned to receive only telephone reminders. The remainder of the sample, 7,567 people, was assigned to two or more treatments. Assignment to the personal canvassing experiment was designed to be uncorrelated with the telephone and mail experiments, so that it could be analyzed separately. Random assignment to each of the telephone/mail treatments was performed in a manner that made calls more frequent among those who received mail. Thus, these two treatments are correlated, and their effects must be estimated using multivariate methods.

Overall, the treatment group for personal canvassing contained 5,794 people, the control group 23,586. For the direct mail experiment, 14,719 people were in the treatment group, and 14,661 were in the control group. The effectiveness of randomization was checked using voter turnout data from the 1996 presidential election. Based on a chi-square test with 15 degrees of freedom for the 16 groups defined in Table 2, we cannot reject the null of independence between treatments and past voting behavior ($p > .10$).

### Personal Canvassing Procedure

During each Saturday and Sunday for four weeks before the election, canvassers were sent to contact randomly selected, registered voters. They were paid $20 per hour and were primarily graduate students. New Haven has a substantial minority population and a significant proportion of non-English speakers. More than half the canvassers were African American or fluent in Spanish, and when possible they were matched to the racial and ethnic composition of the neighborhoods they walked.

For safety reasons, all canvassing was done in pairs and ceased at sunset. This procedure constrained both the pool of available canvassers and our ability to contact people not at home during the day. In contrast to conventional canvassing efforts, we targeted certain households rather than entire streets, which meant that more time was devoted to locating specific addresses and walking from one to the next. Consequently, canvassers were able to contact only 1,615 (28%) of the 5,794 people in the personal canvassing treatment group.4 Examination of the data showed a fairly even

---

3 Random assignment was done at the household level. The results we present treat individuals as the unit of analysis; as we point out below, however, the results are very similar when we look separately at households containing one or two registered voters. Also, the standard errors we report are very similar to the ones obtained using statistical methods that allow for unmodeled similarities between household members, such as generalized least squares or resampling.

4 For the subset of persons not contacted, two supplementary experiments were performed. In certain wards, 719 were randomly chosen to receive a mailer, along with a refrigerator magnet that had the election date printed on it. A separate analysis indicated that this
TABLE 4. Effects of Personal Canvassing on Voter Turnout, by Type of Nonpartisan Appeal

<table>
<thead>
<tr>
<th>Type of Appeal</th>
<th>Turnout Rate</th>
<th>Number of Registered Voters in Treatment Group</th>
<th>Number of Persons Actually Contacted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted Turnout Rates among Experimental Subgroups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civic duty</td>
<td>47.2%</td>
<td>1,985</td>
<td>534</td>
</tr>
<tr>
<td>Neighborhood solidarity</td>
<td>46.3%</td>
<td>1,881</td>
<td>546</td>
</tr>
<tr>
<td>Election is close</td>
<td>48.1%</td>
<td>1,928</td>
<td>535</td>
</tr>
<tr>
<td>Control</td>
<td>44.8%</td>
<td>23,586</td>
<td>N/A</td>
</tr>
<tr>
<td>Implied Effects of Personal Contact on Voter Turnout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civic duty</td>
<td>Turnout Differential (2.43%)/Contact Rate (26.90%) = 9.1% Standard Error (4.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighborhood solidarity</td>
<td>Turnout Differential (1.48%)/Contact Rate (29.03%) = 5.1% Standard Error (4.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Election is close</td>
<td>Turnout Differential (3.36%)/Contact Rate (27.75%) = 12.1% Standard Error (4.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ P_E = \alpha(p_r + t) + (1 - \alpha)p_{nr}, \]  

where the difference between equations 1 and 2 is due to the effect of the experimental treatment. Combining equations 1 and 2, we derive an expression for \( t \): 

\[ t = \frac{P_E - P_C}{\alpha}. \]  

Although the population probabilities are not observed, sample data can be used to obtain an estimate of \( t \). First, using the law of large numbers, 

\[ \text{plim } V_E = P_E, \quad \text{plim } V_C = P_C, \]  

where \( V_E \) is the percentage of the treatment group that votes, and \( V_C \) is the percentage of the control group that votes. Similarly, 

\[ \frac{N_t}{N_E} = \alpha, \]  

where \( N_t \) is the number of subjects in the treatment group who were reached for the experimental treatment, and \( N_E \) is the number of subjects in the treatment group overall. Using equations 4 and 5, we obtain a consistent estimator of \( t \): 

\[ \text{plim } \frac{V_E - V_C}{N_t} = t. \]  

Equation 6 says that, to find the treatment effect, subtract the turnout rate of the control group from the turnout rate of the experimental group and divide this difference by the observed “contact rate,” which is 28%. Using this formula, we find that personal contact raises the probability of turnout by 8.7 percentage points, with a standard error of 2.6. The null hypothesis that canvassing does nothing to increase turnout can be decisively rejected, using a one-tailed test (\( p < .01 \)).

Table 4 suggests that the effects of personal contact do not vary significantly across messages. The close election message boosts turnout rates by 12.1%, which is slightly better than the 9.1% associated with the civic duty appeal and substantially better than the 5.1% for neighborhood solidarity. These findings are suggestive, but the standard errors associated with the estimates are far too large to reject the null hypothesis that the messages have equal effects. Looking ahead to the experiments using direct mail and telephone calls, we find a similar pattern of insignificant differences across messages. Since we cannot rule out the view that any plausible mobilization appeal works equally well, the analysis that follows focuses exclusively on the relative effectiveness of delivering the appeal in person, by telephone, or through the mail.

**Regression Results**

Regression analysis permits us to conduct a more comprehensive analysis, taking into account all the treatments in our experiment. Regression analysis has the further virtue of introducing covariates, such as past voting history, that reduce the unexplained variance in voting rates and allow for more efficient estimation of the experimental effects. For reasons cited above, however, any regression analysis must attend to the possibility that subjects with a higher propensity to vote are easier to reach in person.

Consider the following simple model of how the experimental treatment affects turnout. Suppose again, for purposes of illustration, that the population can be divided into those who are easy to contact and those who

\[ PP = a(pr + pt) + (1 - a)p_{nr}, \]

where

\[ \text{plim } V_E = P_E, \quad \text{plim } V_C = P_C, \]

where \( V_E \) is the percentage of the treatment group that votes, and \( V_C \) is the percentage of the control group that votes. Similarly,

\[ \frac{N_t}{N_E} = \alpha, \]

where \( N_t \) is the number of subjects in the treatment group who were reached for the experimental treatment, and \( N_E \) is the number of subjects in the treatment group overall. Using equations 4 and 5, we obtain a consistent estimator of \( t \):

\[ \text{plim } \frac{V_E - V_C}{N_t} = t. \]

Equation 6 says that, to find the treatment effect, subtract the turnout rate of the control group from the turnout rate of the experimental group and divide this difference by the observed “contact rate,” which is 28%. Using this formula, we find that personal contact raises the probability of turnout by 8.7 percentage points, with a standard error of 2.6. The null hypothesis that canvassing does nothing to increase turnout can be decisively rejected, using a one-tailed test (\( p < .01 \)).

Table 4 suggests that the effects of personal contact do not vary significantly across messages. The close election message boosts turnout rates by 12.1%, which is slightly better than the 9.1% associated with the civic duty appeal and substantially better than the 5.1% for neighborhood solidarity. These findings are suggestive, but the standard errors associated with the estimates are far too large to reject the null hypothesis that the messages have equal effects. Looking ahead to the experiments using direct mail and telephone calls, we find a similar pattern of insignificant differences across messages. Since we cannot rule out the view that any plausible mobilization appeal works equally well, the analysis that follows focuses exclusively on the relative effectiveness of delivering the appeal in person, by telephone, or through the mail.

Regression analysis permits us to conduct a more comprehensive analysis, taking into account all the treatments in our experiment. Regression analysis has the further virtue of introducing covariates, such as past voting history, that reduce the unexplained variance in voting rates and allow for more efficient estimation of the experimental effects. For reasons cited above, however, any regression analysis must attend to the possibility that subjects with a higher propensity to vote are easier to reach in person.

Consider the following simple model of how the experimental treatment affects turnout. Suppose again, for purposes of illustration, that the population can be divided into those who are easy to contact and those who

\[ PP = a(pr + pt) + (1 - a)p_{nr}, \]

where
where \( Y = 1 \) if the subject votes, \( X_1 = 1 \) if the subject is difficult to contact, and \( X_2 = 1 \) if the subject is actually contacted; 0 otherwise. Given that \( X_1 \) is not observed, we might ignore this variable and regress \( Y \) on an intercept and \( X_2 \). This will yield a consistent regression coefficient estimate only if \( X_1 \) and \( X_2 \) are uncorrelated, or if \( b_1 \) equals 0. These special conditions cannot be expected to hold. Unless everyone in the treatment group is contacted, there will be some correlation between how easy it is to reach a subject and the likelihood they are actually reached. It is also quite reasonable to assume that those who are very hard to reach may also be less likely to vote (i.e., \( b_1 \) does not equal 0). Although these points seem straightforward, they have eluded previous research in this area.9

The standard solution to the problem of correlation between a right-hand-side variable and the regression error is to find a suitable instrumental variable. In this case, an ideal instrument is at hand. Recall that a valid instrument satisfies two criteria: The variable must be uncorrelated with the regression error, and it must be correlated with the endogenous variable. The probability that subjects are contacted is a function of whether they are randomly selected for the treatment group. This implies that a dummy variable which equals 1 for subjects in the treatment group will be correlated with the endogenous variable. Because the treatment group is generated through random assignment, there is no reason to suppose that those who are easy to contact will be overrepresented. Thus, the expected correlation between the instrumental variable and the regression error is zero.

Table 5 presents two-stage least-squares regression estimates of the effect of each experimental treatment. As indicated earlier, the instrumental variables used in the regressions indicate whether the person was in a given treatment group. For example, the variable Personal Contact equals 1 if the subject was contacted, and the instrumental variable equals 1 if the person was in the group that we intended to treat. Note that the instrumental variable will be correlated with the included variable (being in the intent-to-treat group predicts the likelihood that one is contacted), but the instrumental variable is not correlated with the regression error (treatment group status is due to random assignment). A similar procedure applies to the telephone experiment, with intent-to-treat serving as an instrument for actual contact. For the mail experiment, the instrumental variable and the independent variable are the same, since the assumed contact rate is 100%.10

Official voting and registration records contain useful information about the sample. For example, we know whether a person voted, abstained, or was absent from the voter rolls in the 1996 general election. We also know an individual’s age, party registration, voting ward, and whether s/he is the sole registered adult in the household or is one of two. Each of these covariates contributes significantly to the predictive accuracy

<table>
<thead>
<tr>
<th>TABLE 5. Linear and Nonlinear Regression of Voter Turnout on Mode of Contact, with and without Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Variables</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Personal contact</td>
</tr>
<tr>
<td>(0.026)</td>
</tr>
<tr>
<td>Direct mailings</td>
</tr>
<tr>
<td>(0 to 3)</td>
</tr>
<tr>
<td>Telephone contact</td>
</tr>
<tr>
<td>(0.023)</td>
</tr>
<tr>
<td>Registered as Democrat</td>
</tr>
<tr>
<td>Republican</td>
</tr>
<tr>
<td>Voted in 1996 general</td>
</tr>
<tr>
<td>election</td>
</tr>
<tr>
<td>Abstained in 1996</td>
</tr>
<tr>
<td>general election</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>(0.0008)</td>
</tr>
<tr>
<td>Age squared</td>
</tr>
<tr>
<td>(0.00007)</td>
</tr>
<tr>
<td>Number of registered</td>
</tr>
<tr>
<td>voters in household</td>
</tr>
<tr>
<td>(1 or 2)</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>Degrees of freedom</td>
</tr>
</tbody>
</table>

Note: The base category for past voting behavior is the set of people who were not registered in 1996. Not reported in this table are the coefficients associated with each of the 29 wards. The first-stage equations include dummy variables representing the intent-to-treat group. This implies that a dummy variable which equals 1 for subjects in the treatment group will be correlated with the endogenous variable. Because the treatment group is generated through random assignment, there is no reason to suppose that those who are easy to contact will be overrepresented. Thus, the expected correlation between the instrumental variable and the regression error is zero.

9 Consider some of the seminal work in this area. Kramer (1970) interprets the higher turnout rate among those reached by a party or candidate as the marginal effect of contact. In the classic study by Eldersveld (1956), those unavailable for personal contact were moved into the control group. This practice results in overestimation of the treatment effect.

10 Our calculations assume that all of the households we intended to treat by mail received the treatment, an assumption implicitly made in all previous mail experiments. In our case, the voter lists were very current and fewer than 1% of the mailings were returned. To adjust the estimated effects for any failure to receive the mail, divide the coefficients in Table 5 by the supposed contact rate.
TABLE 5. Instrumental Variables (IV) Estimates of Average Treatment Effects on Voter Turnout (Percentage Points)

<table>
<thead>
<tr>
<th></th>
<th>Original Data</th>
<th>Revised Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gerber &amp; Green (Incorrect Groups)</td>
<td>Corrected IV (Correct Groups)</td>
</tr>
<tr>
<td>Phone</td>
<td></td>
<td>Phone</td>
</tr>
<tr>
<td>Overall effect</td>
<td>−4.7</td>
<td>−11.6</td>
</tr>
<tr>
<td></td>
<td>(2.3)</td>
<td>(6.6)</td>
</tr>
<tr>
<td>Single-voter households</td>
<td>−13.7</td>
<td>−26.8</td>
</tr>
<tr>
<td></td>
<td>(4.0)</td>
<td>(10.0)</td>
</tr>
<tr>
<td>Two-voter households</td>
<td>1.6</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>(2.7)</td>
<td>(8.7)</td>
</tr>
<tr>
<td>Civic duty</td>
<td>−7.5</td>
<td>−10.8</td>
</tr>
<tr>
<td></td>
<td>(3.0)</td>
<td>(9.9)</td>
</tr>
<tr>
<td>Neighborhood solidarity</td>
<td>−5.1</td>
<td>−6.7</td>
</tr>
<tr>
<td></td>
<td>(4.1)</td>
<td>(4.1)</td>
</tr>
<tr>
<td>Close race</td>
<td>−0.7</td>
<td>−12.4</td>
</tr>
<tr>
<td></td>
<td>(3.5)</td>
<td>(8.3)</td>
</tr>
</tbody>
</table>

Note: The table shows that the negative finding for telephone canvassing is driven by the large and negative effects for single-voter households. The left two columns of estimates display the results based on the incorrectly identified groups used by Gerber and Green (2000). The IV estimates in the middle two columns use the proper treatment assignment and control groups, thereby correcting the original analysis of Gerber and Green (2000). Finally, the estimates in the right columns are based on the revised data using the correct treatment assignment and control groups. Standard errors are in parentheses.

Methods for Evaluating the Implementation of Field Experiments

While the IV method is useful in many situations, the validity of its use relies on the key assumption that treatment assignment is completely randomized. Below, I show that this assumption was violated in Gerber and Green’s experiment and that the violation led to their negative finding about telephone canvassing. Indeed, I now demonstrate, with statistical tests I introduce, that the pattern of incomplete randomization observed in Gerber and Green’s original data would occur with a probability of less than one in 300 million. These results led to the discovery of the implementation errors of their experiment.

The fact that the errors did not occur randomly is another indication of failed randomization in this experiment. For example, Gerber and Green’s revisions of the original data increased the overall rate of compliance for phone calls by five percentage points (see Table 4). This difference is statistically significant (p-value, 0.01), implying that the implementation errors systematically affected those individuals who were more likely to answer the phone when called. Thus, IV estimation, which assumes complete randomization, is not an appropriate method to analyze either the revised or the original data.

Detecting the implementation errors of field experiments is generally a difficult task. The main challenge arises from the fact that statistical tests based on the observed data cannot guarantee that the treatment assignment is randomized with respect to unobserved variables. For this reason, it is advisable to gather as

5. Five percentage points on turnout. Moreover, their inappropriate use of overlapping treatments obscured greater problems. Correcting the treatment assignment and control groups makes the effect even larger, reaching −12 percentage points with a standard error of seven percentage points. These IV estimates based on the original data suggest that get-out-the-vote calls encouraging people to vote discourage them from casting their ballots.

Note that although the negative effect for single-voter households seems to persist in the revised data, the estimated overall effect of phone calls is now small with a large standard error. This is similar to the situation of ITT estimates mentioned above in that the data correction brings Gerber and Green’s estimates closer to positive effects. As I show below, however, data correction alone is not sufficient to fix the implementation errors.

Finally, the corrected IV estimates for personal visits are much greater than those from the original analysis for both original and revised data, reaching to an increase of more than 10 percentage points in turnout. This significant difference is solely due to the correction of treatment assignment and control groups. This is clear evidence against the assumption of Gerber and Green (2000, 660) that the effects of different canvassing methods are constant and additive.

While Gerber and Green’s two-stage least-squares analysis (with all covariates) indicates a smaller negative effect, their two-stage probit analysis shows that the effect of phone calls is about negative five percentage points and statistically significant.
• Problems with Implementation

  – Survey company got lists confused with another of Gerber and Green’s experiments

  – Sent wrong survey (about blood donation) to subjects

• Kosuke Imai shows lack of balance in pre-existing covariates

\[ T_i = f \left( \overrightarrow{X} \beta \right) + \epsilon_i \]

where \( T_i \) is a vector of treatment variables, \( f \) is the logistic function and \( \overrightarrow{X} \) is a vector of supposedly balanced covariates. Then the author does a joint test of the vector \( \beta \) :

\[ H_0 : \beta = 0 \]

• Imai then uses matching methods to estimate treatment effects within covariate-balanced groups.
candidates may want to know about how many visits or postcards are necessary to increase voter turnout by one percentage point. In this case, it is not necessary to know how many voters actually talked to canvassers or read postcards. On the other hand, political scientists, who want to assess the relative effectiveness of various canvassing methods need this extra information. Even when personal canvassing seems less effective, for example, it may only appear ineffective because voters are more difficult to reach by visits than by postcards. Hence, the different compliance rates for the two methods become critical.

THE NEW HAVEN VOTER MOBILIZATION STUDY

In this section, I replicate and extend Gerber and Green’s analysis of the voter mobilization study. Gerber and Green (2000) designed and conducted an experiment where registered voters in randomly selected households of New Haven were encouraged to vote in the 1998 general election by means of personal visits, phone calls, and postcards. They then examined voting records and analyzed which strategies had increased voter turnout. In addition to the voting record of the 1998 election, the data include covariates that describe the following characteristics of each registered voter: number of registered voters in the household (one or two), age, party affiliation (registered Democrats, registered Republicans, or others), voting record in the last general election (voted, did not vote, or was not registered for 1996 election), and ward of residence in New Haven (29 wards).

Inefficient Experimental Design

Table 1 shows the unusually complicated experimental design of the original study with the substantial overlap of different treatment assignments. Over 40% of voters in the sample were assigned more than one treatment. For example, 122 voters were assigned to receive three postcards, a phone call, and a personal visit with the civic duty message. Further variation in the nature of the treatment was possible because Gerber and Green used three different appeal messages: civic duty, neighborhood solidarity, and close election. The authors note that the neighborhood solidarity message was not used for phone calls (Gerber and Green 2000, 656). Altogether, this design produced a total of 45 different treatment combinations and their corresponding potential outcomes.

Such complex experimental design leads to the inefficient estimation of treatment effects unless one makes arbitrary assumptions. This is unfortunate since the advantage of experimental methods is to avoid additional assumptions that are often necessary in observational studies. For example, Gerber and Green (2000) assume that the effect of telephone canvassing remains the same regardless of whether voters have received other treatments. However, phone calls may not increase the probability of voting as much for those voters who already have received a personal visit. Furthermore, the timing of contact differs from one canvassing method to another and this variation was not randomized; e.g., phone calls were made during the three days prior to the election, whereas personal visits were made over a period of four weeks. Such systematic differences in the administration of multiple treatments will yield incorrect inferences unless properly controlled in the analysis.

Incorrectly Identified Treatment Assignment and Control Groups

Gerber and Green (2000) also incorrectly identified the treatment assignment and control groups used in their field experiment and, as such, failed to estimate their causal quantities of interest. For example, when estimating the marginal effect of phone calls, Gerber and Green used the treatment assignment group that includes those who were also assigned other treatments such as personal visits and postcards (the upper two rows in Table 1). Their control group included those voters who were assigned other treatments (all categories in the bottom two rows in Table 1). In order to correctly estimate the treatment and ITT effects, the appropriate control group should consist solely of the 10,800 voters who were assigned no treatment and hence received no intervention. Likewise, the members of the treatment assignment group for phone calls should not include those who were assigned any other treatment.

| TABLE 1. The Original Experimental Design Reported in Gerber and Green (2000) |
|-----------------|----------------|------------|-----------|-----------|
| Phone Visit     | None | Once | Twice | 3 times |
| Civic            | 33   | 103  | 126   | 122      |
| Neighbor/civic  | 74   | 144  | 113   | 127      |
| Close           | 110  | 138  | 113   | 134      |
| No visit        | 581  | 443  | 432   | 479      |
| Civic           | 0    | 491  | 520   | 542      |
| Neighbor/civic | 377  | 517  | 534   | 501      |
| Close          |       |      |       |          |

Note: The figures represent the number of registered voters in New Haven for each treatment assignment combination. For example, 122 voters were assigned to receive three postcards, a phone call, and a personal visit with the civic duty message. Treatment assignment groups of interest are underlined. A box highlights the large control group.

*For phone calls, the civic duty appeal was used instead of the neighborhood solidarity message (Gerber and Green 2000, 656).*
### TABLE 2. Treatment Assignment and Control Groups Based on the Revised Data

<table>
<thead>
<tr>
<th>Mail</th>
<th>None</th>
<th>Once</th>
<th>Twice</th>
<th>3 times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civic</td>
<td>0</td>
<td>88</td>
<td>107</td>
<td>98</td>
</tr>
<tr>
<td>Civic/blood</td>
<td>104</td>
<td>17</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>Civic/blood-civic</td>
<td>0</td>
<td>12</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Neighbor</td>
<td>0</td>
<td>109</td>
<td>92</td>
<td>101</td>
</tr>
<tr>
<td>Neighbor/civic</td>
<td>74</td>
<td>22</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Neighbor/civic-neighbor</td>
<td>0</td>
<td>13</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Close</td>
<td>110</td>
<td>138</td>
<td>113</td>
<td>134</td>
</tr>
<tr>
<td>No visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civic</td>
<td>428</td>
<td>385</td>
<td>352</td>
<td>411</td>
</tr>
<tr>
<td>Civic/blood</td>
<td>371</td>
<td>84</td>
<td>98</td>
<td>95</td>
</tr>
<tr>
<td>Civic/blood-civic</td>
<td>0</td>
<td>29</td>
<td>46</td>
<td>33</td>
</tr>
<tr>
<td>Neighbor</td>
<td>0</td>
<td>374</td>
<td>367</td>
<td>390</td>
</tr>
<tr>
<td>Neighbor/civic</td>
<td>0</td>
<td>73</td>
<td>102</td>
<td>97</td>
</tr>
<tr>
<td>Neighbor/civic-neighbor</td>
<td>0</td>
<td>44</td>
<td>51</td>
<td>55</td>
</tr>
<tr>
<td>Close</td>
<td>377</td>
<td>517</td>
<td>534</td>
<td>501</td>
</tr>
<tr>
<td>No phone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civic</td>
<td>940</td>
<td>136</td>
<td>202</td>
<td>216</td>
</tr>
<tr>
<td>Neighbor</td>
<td>853</td>
<td>175</td>
<td>201</td>
<td>194</td>
</tr>
<tr>
<td>Close</td>
<td>822</td>
<td>194</td>
<td>211</td>
<td>206</td>
</tr>
<tr>
<td>No visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civic</td>
<td>815</td>
<td>858</td>
<td>765</td>
<td></td>
</tr>
<tr>
<td>Neighbor</td>
<td>10,582</td>
<td>764</td>
<td>849</td>
<td>767</td>
</tr>
<tr>
<td>Close</td>
<td>772</td>
<td>817</td>
<td>783</td>
<td></td>
</tr>
</tbody>
</table>

Note: The figures represent the number of registered voters in New Haven for each treatment assignment combination. For example, 104 voters were assigned a phone call with the blood donation message and a personal visit with the civic duty appeal. Treatment assignment groups of interest are underlined. A box highlights the control group.

* For phone calls, the blood donation appeal was used instead of the civic duty message.
* For phone calls, either the blood donation or the civic duty appeal was used.
* For phone calls, the civic duty appeal was used instead of the neighborhood solidarity message.
* For phone calls, either the civic duty or the neighborhood solidarity appeal was used.

This implies that the ITT and treatment effects reported in Gerber and Green (2000) are confounded by the effects of other treatments. In experiments, an appropriate control group is critical to ensure internal validity (e.g., Campbell and Stanley 1963). In principle, it is advisable to minimize the number of treatments in field experiments. Although factorial designs may be feasible in laboratory experiments, additional complications such as noncompliance make it difficult to estimate the effects of multiple overlapping treatments in field experiments. In this article, I focus on the marginal effects of each treatment rather than their interaction effect, as the latter would involve additional assumptions and few data are available to estimate such quantities.

As noted above, the analysis in the initial draft of this article detected the implementation errors and led to the subsequent revisions of the original data. Table 2 shows the treatment assignment and control groups based on the most recent data and Gerber and Green’s latest version of their experimental design. The total number of treatment combinations is now seventy, making the experimental design even more complex. For the analysis of the revised data, I correct the treatment group for telephone canvassing to include only those voters who were assigned no other treatment. I also exclude those who were possibly assigned the blood donation messages. This yields the total of 428 voters with the civic duty appeal and 377 individuals with the close race message. The new control group consists of 10,582 voters who were assigned no treatment.

The analysis of the revised data reveals discrepancies between Gerber and Green’s description of the implementation errors and the altered coding scheme.
TABLE 3. Estimated Average Intention-To-Treat (ITT) Effects on Voter Turnout Assuming Complete Randomization (Percentage Points)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Gerber &amp; Green (Incorrect Groups)</th>
<th>Corrected ITT (Correct Groups)</th>
<th>Revised Data (Correct Groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone(^a)</td>
<td>−1.5 (0.7)</td>
<td>−2.9 (1.7)</td>
<td>−0.9 (1.8)</td>
</tr>
<tr>
<td>Visit</td>
<td>2.4 (0.7)</td>
<td>3.9 (1.1)</td>
<td>3.6 (1.1)</td>
</tr>
<tr>
<td>Mail</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once</td>
<td>0.6 (0.3)</td>
<td>0.4 (1.1)</td>
<td>0.5 (1.1)</td>
</tr>
<tr>
<td>Twice</td>
<td>1.2 (0.5)</td>
<td>0.8 (1.1)</td>
<td>0.8 (1.1)</td>
</tr>
<tr>
<td>3 times</td>
<td>1.7 (0.8)</td>
<td>2.6 (1.1)</td>
<td>2.7 (1.1)</td>
</tr>
</tbody>
</table>

Note: The left column of estimates displays the results based on the incorrectly identified groups as published in Gerber and Green (2000). The ITT estimates in the middle column use the proper treatment assignment and control groups, thereby correcting the original analysis of Gerber and Green (2000). Finally, the estimates in the right column are based on the revised data using the correct treatment assignment and control groups. Standard errors are in parentheses.\(^a\) The ITT effect of phone calls was not reported by Gerber and Green (2000) and is calculated based on their method.

For example, on their Web site they describe one of their errors as follows: “Subjects who would have received Civic Duty mail or personal appeals received phone appeals requesting a Blood Donation” (see footnote 4). Although this error should not affect the control group of those who were assigned no treatment in the first place, the revised control group has about 300 voters fewer than the original group. Such remaining inconsistency calls for further clarifications about the coding changes beyond what is currently documented.

ANALYSIS ASSUMING COMPLETE RANDOMIZATION WITH CORRECTED TREATMENT ASSIGNMENT AND CONTROL GROUPS

With the corrected treatment assignment and control groups, I reestimate the average ITT and treatment effects by applying the statistical method used in Gerber and Green (2000), which assumes complete randomization of treatment assignments.

Estimation of the ITT Effect

Under the assumption of complete randomization, the treatment assignment is independent of all observed and unobserved individual characteristics. Therefore, the difference in the sample means of the treatment assignment and control groups is an unbiased estimate of the average ITT effect. Namely,

\[
\hat{\text{ITT}} = \frac{\sum_{i=1}^{N} Y_i Z_i}{N_1} - \frac{\sum_{i=1}^{N} Y_i (1 - Z_i)}{N_0},
\]  

where \(N_1 = \sum_{i=1}^{N} Z_i\) is the size of the treatment assignment group, \(N_0 = \sum_{i=1}^{N} (1 - Z_i)\) is the size of the control group, and \(N = N_0 + N_1\).\(^6\)

Table 3 shows the results of the ITT analysis using the correct treatment and control groups. First, the corrected ITT analysis in the middle column confirms the conclusion of Gerber and Green (2000) that personal canvassing is the most effective method for increasing voter turnout. Second, get-out-the-vote calls have a significant negative effect on turnout. Using the appropriate treatment assignment and control groups does not change the odd finding of the original article that telephone canvassing reduces voter turnout.

As one would expect, altering the data also changes the estimates. The analysis of the revised data with correct groups (in the right column) suggests that the overall ITT effect of phone calls is only slightly negative, with a larger standard error. In the next section, however, I show that the data correction alone does not solve the entire problem. In principle, the implementation errors of field experiments cannot be fixed by the experimenter after the fact without statistical adjustments.

Mail canvassing also mobilizes voters. (Gerber and Green 2000, 661) argued that “even if the effective marginal costs of canvassing were doubled, face-to-face mobilization would still be cost effective.” This conclusion, however, is based on their assumption that all voters who were sent postcards actually received and read them (659, fn 10). Such an assumption is not warranted because many cards may not have reached a voter due to changes of address or may have been

\(^6\) In the case of phone calls, for example, \(N_1 = 958\) and \(N_0 = 10,800\).
randomization observed in Gerber and Green’s data can occur only with a probability of one in 300 million. This probability is smaller for the revised data, reaching to one in 2 billion. (Note that a small sample size makes it harder to detect failure of randomization, so that the larger $p$-value for phone calls than for visits and mailings does not necessarily imply that randomization was more successful.) In sum, the test with respect to observed covariates also provides strong evidence that treatment assignment was not randomized in Gerber and Green’s field experiment.

In field experiments, randomization of treatment assignment is not as easy to accomplish as one might expect. In practice, it is often difficult to randomize every aspect of each treatment. In Gerber and Green’s experiment, personal canvassing was conducted over a period of four weeks before the election, whereas telephone canvassing took place over three days including the election day. Postcards were sent out during the two weeks before the election. Although a visit right before the election would have a greater effect than a visit one month before the election day, the timing of contact was not randomized. Likewise, the effect of different canvassers, if not randomized, can confound the effect of different canvassing methods. These examples illustrate the difficulty of randomization and potential confounding effects that threaten the validity of field experiments.

Finally, I investigate the sources of the negative finding about phone calls. Both Gerber and Green’s analysis and the corrected IV analysis indicate that telephone canvassing has a large and negative effect on voter turnout among single-voter households. I find that for this subgroup the assignment of phone calls was not randomized with respect to the past voting record. In particular, only 42% of the treatment assignment group voted in the last election, whereas 47% of the control group voted ($p$-value, 0.05). The randomization for this group appears to be incomplete even with the incorrectly identified treatment assignment and control groups used by Gerber and Green. Since those who voted in the last election are 40 percentage points more likely to vote in the current election on average, this difference contributes to the large negative effects of phone calls for single-voter households.

### When One Should Not Use IV Estimation

The large bias of IV estimation that results from violation of the exclusion restriction is well documented (e.g., Angrist, Imbens, and Rubin 1996, 450). In particular, the bias is worsened when unbalanced variables are good predictors of the outcome variable and when a large number of noncompliers exist. Equation (4) illustrates these two conditions; the bias of the IV estimate is large (a) when the bias of the ITT estimate due to incomplete randomization is large and (b) when the compliance rate is low. (Recall that the IV estimate is equal to the ITT estimate divided by the estimated compliance rate.)

Gerber and Green’s study fits both conditions for large bias. First, the unbalanced covariates (i.e., the voting record in the previous election) predict turnout well, which suggests that the bias in the estimated ITT effect is large. Furthermore, the compliance rate of this field experiment is low (about 25% for phone calls). This low compliance rate implies that if the ITT effect is biased by five percentage points, for example, then the bias of the IV estimate can be as large as 20 percentage points. Thus, the combination of a large bias in the ITT estimate and low compliance rate led to the puzzling finding that get-out-the-vote calls significantly decrease turnout.

If one successfully randomizes the treatment assignment, the method of instrumental variables can give estimated treatment effects that are consistent in large samples. However, as the analysis of this section suggests, making this assumption in practice requires careful experimental design and successful implementation. In this case, the failure of randomization for telephone canvassing led to inaccurate causal inferences.

---

**TABLE 6. Probability of Successful Randomization with Respect to Observed Covariates in Gerber and Green’s Field Experiment**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Original Data</th>
<th>Revised Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Probability</td>
<td>$N$</td>
</tr>
<tr>
<td>Phone</td>
<td>0.035</td>
<td>958</td>
</tr>
<tr>
<td>Visit</td>
<td>0.000012</td>
<td>2,686</td>
</tr>
<tr>
<td>Mail</td>
<td>0.0000000035</td>
<td>7,369</td>
</tr>
</tbody>
</table>

Note: Probability represents the $p$-value of the residual deviance test from a logistic regression model predicting the assignment of each treatment given all observed covariates and their first-order interactions. $N$ represents the size of the treatment assignment group. The last row in the second column, for example, tells us that under the assumption of successful randomization, the pattern of incomplete randomization for mailings observed in Gerber and Green’s original data would occur only with a probability of about one in 300 million. These probabilities cannot be compared across different treatments because of different sample sizes.

---

17 Compared with the control group, the treatment assignment group includes significantly more individuals who abstained in the last election. The mean difference is statistically significant at the 0.05 level.

18 It is also important to note the finite sample bias and inefficiency of IV estimation (e.g., Bound, Jaeger, and Baker 1995). The small size of each treatment group in the New Haven mobilization study suggests the importance of finite sample consideration.
Gerber, Karlan and Bergan I

• Idea: Randomize Access to Newspapers of Different Biases to See Effect of Media Bias on Voting, Knowledge Preferences:

• The problem with just looking at voting patterns on newspaper reading:

\[ V_j = \sigma + \lambda N_j + \pi_j \]

• Is that both are determined by a mutual variable ideology:

\[ V_j = \alpha + \beta I_j + \varepsilon_j \]

\[ N_j = \mu + \gamma I_j + \varepsilon_j \]
Gerber, Karlan and Bergan II

• In the case where we run the naïve regression of voting on newspaper reading then (even if there is no direct effect of newspaper reading on voting), we get:

\[ \lambda = \frac{\text{cov}(V_j, N_j)}{\text{var}(N_j)} = \frac{\beta \gamma \sigma_i^2}{\sigma_N^2} \]

• Gerber, Karlan and Bergan randomize access to newspapers to solve this problem of selection based on ideology.
Gerber, Karlan and Bergan III

- Gerber, Karlan and Bergan did a baseline survey in September, 2005 (3347 responses, 1065 in follow up)
- In Prince William County, Virginia, 25 miles from Washington, DC
- Randomly selected participants from consumer database (46%) and voter registration database (54%)
- Dropped
  - people already subscribing to one of the two newspapers (the Washington Post and the Washington Times)
  - people who did not answer at least one question on the baseline survey
Gerber, Karlan and Bergan IV

• Randomized into Washington Post, Washington Times, and Control based upon stratification by answers to question on:
  – Subscription to a magazine
  – Subscription to a non-Post and non-Times newspaper
  – Who they planned to vote for
  – Whether they said they wanted to read more news
Gerber, Karlan and Bergan IV

- Double Coverage: 75 of those who participated already were receiving the Washington Post (maybe just Sunday), 5 already received the Washington Times

- Non-delivery: 76 households did not receive the Times because they were not in the delivery area, 1 in the Post area

- Attrition: 59 (out of 965) dropped their Post subscription, 54 (out of 950) dropped their Times subscription

- Attrition: 306 Post subscribers responded to the follow, 313 Times, and 446 control (1087 out of 3347 responded)
<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Wave 1</th>
<th>Wave 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post</td>
<td>605</td>
<td>360</td>
<td>965</td>
</tr>
<tr>
<td></td>
<td>28.8</td>
<td>29.0</td>
<td>28.9</td>
</tr>
<tr>
<td>Times</td>
<td>595</td>
<td>355</td>
<td>950</td>
</tr>
<tr>
<td></td>
<td>28.3</td>
<td>28.6</td>
<td>28.4</td>
</tr>
<tr>
<td>Control</td>
<td>904</td>
<td>528</td>
<td>1,432</td>
</tr>
<tr>
<td></td>
<td>43.0</td>
<td>42.5</td>
<td>42.8</td>
</tr>
<tr>
<td>Totals</td>
<td>2,104</td>
<td>1,243</td>
<td>3,347</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: Cell entries indicate number of individuals assigned to each treatment group. Numbers in italics are column percentages.
### Table 2A: Summary Statistics from Baseline Survey
**Mean and standard errors**

#### Panel A: Baseline Survey Responses

<table>
<thead>
<tr>
<th>Sample Average</th>
<th>Control</th>
<th>Post</th>
<th>Times</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>% female</td>
<td>34.8</td>
<td>34.4</td>
<td>33.0</td>
<td>37.0</td>
</tr>
<tr>
<td></td>
<td>(0.8)</td>
<td>(1.3)</td>
<td>(1.5)</td>
<td>(1.6)</td>
</tr>
<tr>
<td>% voted in 2004</td>
<td>88.6</td>
<td>88.5</td>
<td>88.8</td>
<td>88.6</td>
</tr>
<tr>
<td></td>
<td>(0.8)</td>
<td>(1.2)</td>
<td>(1.4)</td>
<td>(1.4)</td>
</tr>
<tr>
<td>% voted in 2002</td>
<td>48.0</td>
<td>49.0</td>
<td>45.8</td>
<td>49.1</td>
</tr>
<tr>
<td></td>
<td>(1.2)</td>
<td>(1.9)</td>
<td>(2.3)</td>
<td>(2.3)</td>
</tr>
<tr>
<td>% voted in 2001</td>
<td>7.3</td>
<td>7.1</td>
<td>7.7</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(1.0)</td>
<td>(1.2)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>% from consumer list</td>
<td>50.9</td>
<td>52.6</td>
<td>50.0</td>
<td>49.3</td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(1.3)</td>
<td>(1.6)</td>
<td>(1.6)</td>
</tr>
<tr>
<td>% get news or political magazine</td>
<td>9.2</td>
<td>9.4</td>
<td>8.8</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(0.8)</td>
<td>(0.9)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>% prefers Democratic candidate for Governor in VA</td>
<td>14.4</td>
<td>14.5</td>
<td>14.6</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(0.9)</td>
<td>(1.1)</td>
<td>(1.1)</td>
</tr>
<tr>
<td>% no preference in VA Gov. race</td>
<td>14.8</td>
<td>14.2</td>
<td>15.5</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(0.9)</td>
<td>(1.2)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>% in wave 2 of random assignment</td>
<td>37.1</td>
<td>36.9</td>
<td>37.3</td>
<td>37.4</td>
</tr>
<tr>
<td></td>
<td>(0.8)</td>
<td>(1.3)</td>
<td>(1.6)</td>
<td>(1.6)</td>
</tr>
<tr>
<td>% participating in follow-up</td>
<td>32.3</td>
<td>31.7</td>
<td>32.0</td>
<td>33.5</td>
</tr>
<tr>
<td></td>
<td>(0.8)</td>
<td>(1.2)</td>
<td>(1.5)</td>
<td>(1.5)</td>
</tr>
<tr>
<td>N</td>
<td>3347</td>
<td>1432</td>
<td>965</td>
<td>950</td>
</tr>
</tbody>
</table>

#### Panel B: Baseline Survey Responses on the Sample of Those Who Completed the Follow-up Survey

<table>
<thead>
<tr>
<th>Sample Average</th>
<th>Control</th>
<th>Post</th>
<th>Times</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>% female</td>
<td>32.9</td>
<td>31.5</td>
<td>36.8</td>
<td>30.9</td>
</tr>
<tr>
<td></td>
<td>(1.5)</td>
<td>(2.2)</td>
<td>(2.8)</td>
<td>(2.6)</td>
</tr>
<tr>
<td>% voted in 2004</td>
<td>90.7</td>
<td>92.6</td>
<td>89.2</td>
<td>89.5</td>
</tr>
<tr>
<td></td>
<td>(1.2)</td>
<td>(1.7)</td>
<td>(2.5)</td>
<td>(2.3)</td>
</tr>
<tr>
<td>% voted in 2002</td>
<td>56.0</td>
<td>57.6</td>
<td>50.6</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>(2.1)</td>
<td>(3.3)</td>
<td>(4.0)</td>
<td>(3.8)</td>
</tr>
<tr>
<td>% voted in 2001</td>
<td>8.4</td>
<td>9.2</td>
<td>8.2</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>(1.2)</td>
<td>(1.8)</td>
<td>(2.2)</td>
<td>(2.0)</td>
</tr>
<tr>
<td>% from consumer list</td>
<td>48.3</td>
<td>49.6</td>
<td>48.9</td>
<td>45.9</td>
</tr>
<tr>
<td></td>
<td>(1.5)</td>
<td>(2.4)</td>
<td>(2.9)</td>
<td>(2.8)</td>
</tr>
<tr>
<td>% get news or political magazine</td>
<td>11.3</td>
<td>10.4</td>
<td>11.0</td>
<td>12.9</td>
</tr>
<tr>
<td></td>
<td>(1.0)</td>
<td>(1.4)</td>
<td>(1.8)</td>
<td>(1.9)</td>
</tr>
<tr>
<td>% prefers Democratic candidate for Governor in VA</td>
<td>19.4</td>
<td>19.6</td>
<td>21.0</td>
<td>17.6</td>
</tr>
<tr>
<td></td>
<td>(1.2)</td>
<td>(1.9)</td>
<td>(2.3)</td>
<td>(2.2)</td>
</tr>
<tr>
<td>% no preference in VA Gov. race</td>
<td>12.9</td>
<td>13.2</td>
<td>10.0</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>(1.0)</td>
<td>(1.6)</td>
<td>(1.7)</td>
<td>(2.0)</td>
</tr>
<tr>
<td>% in wave 2 of random assignment</td>
<td>35.1</td>
<td>35.0</td>
<td>38.5</td>
<td>31.8</td>
</tr>
<tr>
<td></td>
<td>(1.5)</td>
<td>(2.3)</td>
<td>(2.8)</td>
<td>(2.6)</td>
</tr>
<tr>
<td>N</td>
<td>1,065</td>
<td>446</td>
<td>306</td>
<td>313</td>
</tr>
</tbody>
</table>

Note: Standard errors reported in parentheses. Column 5 reports the p-values for chi squared tests of independence between treatments for each baseline variable.
Table 2B: Summary Statistics for Outcome Measures
Mean and Standard Errors

<table>
<thead>
<tr>
<th></th>
<th>Sample Avg.</th>
<th>Control</th>
<th>Post</th>
<th>Times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Voted</td>
<td>.728</td>
<td>.726</td>
<td>.725</td>
<td>.735</td>
</tr>
<tr>
<td></td>
<td>(.014)</td>
<td>(.021)</td>
<td>(.025)</td>
<td>(.025)</td>
</tr>
<tr>
<td>Voted for Democrat</td>
<td>.446</td>
<td>.411</td>
<td>.490</td>
<td>.451</td>
</tr>
<tr>
<td></td>
<td>(.019)</td>
<td>(.029)</td>
<td>(.035)</td>
<td>(.034)</td>
</tr>
<tr>
<td>Did not Vote, But Preferred Democrat</td>
<td>.399</td>
<td>.419</td>
<td>.416</td>
<td>.351</td>
</tr>
<tr>
<td></td>
<td>(.030)</td>
<td>(.046)</td>
<td>(.056)</td>
<td>(.055)</td>
</tr>
<tr>
<td>Voted for or Preferred Democrat</td>
<td>.433</td>
<td>.413</td>
<td>.470</td>
<td>.425</td>
</tr>
<tr>
<td></td>
<td>(.016)</td>
<td>(.024)</td>
<td>(.030)</td>
<td>(.029)</td>
</tr>
<tr>
<td>Most important Problem (1=issue other than scandals, 0=scandals)</td>
<td>.078</td>
<td>.08</td>
<td>.068</td>
<td>.086</td>
</tr>
<tr>
<td></td>
<td>(.008)</td>
<td>(.013)</td>
<td>(.014)</td>
<td>(.016)</td>
</tr>
<tr>
<td>Most important issues in Iraq (1=constitution or Hussein trial)</td>
<td>.444</td>
<td>.442</td>
<td>.472</td>
<td>.417</td>
</tr>
<tr>
<td></td>
<td>(.015)</td>
<td>(.024)</td>
<td>(.029)</td>
<td>(.028)</td>
</tr>
<tr>
<td>Leak case (3=no one did anything wrong; 1=something illegal)</td>
<td>1.75</td>
<td>1.74</td>
<td>1.72</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.038)</td>
<td>(.047)</td>
<td>(.045)</td>
</tr>
<tr>
<td>Alito confirmation (3=should confirm, 1=should not confirm)</td>
<td>2.34</td>
<td>2.37</td>
<td>2.27</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td>(.021)</td>
<td>(.033)</td>
<td>(.040)</td>
<td>(.037)</td>
</tr>
<tr>
<td>Specific issue index (higher scores conservative)</td>
<td>.021</td>
<td>.033</td>
<td>-.028</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>(.020)</td>
<td>(.032)</td>
<td>(.039)</td>
<td>(.035)</td>
</tr>
<tr>
<td>Bush Approval (4=strong approval, 1=strong disapproval)</td>
<td>2.43</td>
<td>2.48</td>
<td>2.37</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td>(.043)</td>
<td>(.066)</td>
<td>(.079)</td>
<td>(.081)</td>
</tr>
<tr>
<td>Republican favorable (4=very favorable, 1=very unfavorable)</td>
<td>1.47</td>
<td>1.50</td>
<td>1.41</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>(.032)</td>
<td>(.050)</td>
<td>(.058)</td>
<td>(.059)</td>
</tr>
<tr>
<td>Conservatism (7=extreme conservative, 1=extreme liberal)</td>
<td>4.51</td>
<td>4.56</td>
<td>4.38</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td>(.045)</td>
<td>(.069)</td>
<td>(.087)</td>
<td>(.083)</td>
</tr>
<tr>
<td>Broad policy index</td>
<td>.001</td>
<td>.038</td>
<td>-.066</td>
<td>.014</td>
</tr>
<tr>
<td></td>
<td>(.025)</td>
<td>(.039)</td>
<td>(.046)</td>
<td>(.047)</td>
</tr>
<tr>
<td>Broad and specific issue index</td>
<td>.010</td>
<td>.033</td>
<td>-.046</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td>(.021)</td>
<td>(.032)</td>
<td>(.038)</td>
<td>(.038)</td>
</tr>
<tr>
<td>Knew number dead in Iraq</td>
<td>.784</td>
<td>.781</td>
<td>.779</td>
<td>.791</td>
</tr>
<tr>
<td></td>
<td>(.013)</td>
<td>(.019)</td>
<td>(.024)</td>
<td>(.023)</td>
</tr>
<tr>
<td>Identified Libby as involved in leak</td>
<td>.739</td>
<td>.754</td>
<td>.705</td>
<td>.748</td>
</tr>
<tr>
<td></td>
<td>(.013)</td>
<td>(.020)</td>
<td>(.026)</td>
<td>(.025)</td>
</tr>
<tr>
<td>Identified Miers as Supreme Court nominee</td>
<td>.777</td>
<td>.785</td>
<td>.729</td>
<td>.813</td>
</tr>
<tr>
<td></td>
<td>(.013)</td>
<td>(.019)</td>
<td>(.026)</td>
<td>(.022)</td>
</tr>
<tr>
<td>Fact index</td>
<td>-.009</td>
<td>.007</td>
<td>-.079</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td>(.022)</td>
<td>(.034)</td>
<td>(.043)</td>
<td>(.040)</td>
</tr>
<tr>
<td>N</td>
<td>1065</td>
<td>446</td>
<td>306</td>
<td>313</td>
</tr>
</tbody>
</table>
Table 4: Effect of Post or Times on Voting Behavior in Virginia Governors Race

<table>
<thead>
<tr>
<th></th>
<th>Voted</th>
<th>Voted for Democrat</th>
<th>Did not vote, but Preferred Democrat</th>
<th>Voted for or Preferred Democrat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1a)</td>
<td>(1b)</td>
<td>(1c)</td>
<td>(2a)</td>
</tr>
<tr>
<td>Post</td>
<td>-0.001</td>
<td>0.018</td>
<td>-0.008</td>
<td>0.079*</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.032)</td>
<td>(0.034)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Times</td>
<td>0.009</td>
<td>0.026</td>
<td>0.012</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.031)</td>
<td>(0.034)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>N</td>
<td>1079</td>
<td>1040</td>
<td>1040</td>
<td>718</td>
</tr>
<tr>
<td>Refused</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td>Does not know</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing Cov.</td>
<td>0</td>
<td>39</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>Not asked</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>294</td>
</tr>
<tr>
<td>Total Surveyed</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.00</td>
<td>0.34</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Covariates</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Strata indicators</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Surveyor/Date indicators</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. ** 95% significance, * 90% significance. Dependent variables in the four sets of columns are as follows: self-reported voter turnout, voted for the Democratic candidate (among those who claimed to vote), preferred the Democrat (among those who did not vote), and either preferred the Democratic candidate (if they said they did not vote) or voted for the Democratic candidate (if they said they voted), respectively. In the row labeled “covariates”, we refer to data from the baseline survey: gender, reported age, three separate indicators for voting in the 2001, 2002 and 2004 general elections, an indicator for whether the respondent was drawn from a consumer list, self report of receiving any news or political magazines, and baseline survey self reports of preferring the Republican candidate in the gubernatorial election and having no preference in the gubernatorial election, and an indicator for wave of the study. In the row “strata indicators”, we include indicator variables for each strata formed prior to the randomization, which included unique combinations of the following: intention to vote, receive a paper (non-Post/non-Times), mentions ever reading a paper, gets a magazine, and asked whether they wish they read the paper more. “Surveyor/Date indicators” refers to a set of indicator variables for each unique combination of surveyor and date for the follow-up survey. All results remain qualitatively similar, and statistical significance remains as-is, using probit or ordered probit specifications instead of OLS.
Table 5: The Effect of Treatment on Attitudes Towards National Politics

**Panel A: Specific Issues**

<table>
<thead>
<tr>
<th></th>
<th>Most important problem (1=issue other than scandals, 0=scandals)</th>
<th>Most important issues in Iraq (1=constitution or Hussein trial)</th>
<th>Leak case (3=no one did anything wrong; 1=something illegal)</th>
<th>Alito confirmation (3=should confirm, 1=should not confirm)</th>
<th>Specific Issue Index (higher scores conservative)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1a)</td>
<td>(1b)</td>
<td>(1c)</td>
<td>(2a)</td>
<td>(2b)</td>
</tr>
<tr>
<td>Post</td>
<td>-.012</td>
<td>-.021</td>
<td>-.028</td>
<td>.038</td>
<td>.020</td>
</tr>
<tr>
<td>Times</td>
<td>.005</td>
<td>.013</td>
<td>.013</td>
<td>-.020</td>
<td>-.004</td>
</tr>
<tr>
<td>N</td>
<td>1033</td>
<td>996</td>
<td>996</td>
<td>982</td>
<td>949</td>
</tr>
<tr>
<td>Refused</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>DK</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Missing Cov.</td>
<td>0</td>
<td>37</td>
<td>37</td>
<td>0</td>
<td>67</td>
</tr>
<tr>
<td>Total Surveyed</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
</tr>
<tr>
<td>R-squared</td>
<td>.00</td>
<td>.14</td>
<td>.24</td>
<td>.00</td>
<td>.30</td>
</tr>
<tr>
<td>Covariates?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Strata indicators?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Operator / date indicators?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**Panel B: Broad National Issues**

<table>
<thead>
<tr>
<th></th>
<th>Bush Approval Rating (4=strong approval, 1=strong disapproval)</th>
<th>Republican Favorable (4=very favorable, 1=very unfavorable)</th>
<th>Conservatism (7=extreme conservative, 1=extreme liberal)</th>
<th>Broad policy Index</th>
<th>Broad and Specific Issue Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(6a)</td>
<td>(6b)</td>
<td>(6c)</td>
<td>(7a)</td>
<td>(7b)</td>
</tr>
<tr>
<td>Post</td>
<td>-.114</td>
<td>-.046</td>
<td>-.164</td>
<td>-.096</td>
<td>-.015</td>
</tr>
<tr>
<td>Times</td>
<td>-.058</td>
<td>-.056</td>
<td>-.165</td>
<td>-.026</td>
<td>-.010</td>
</tr>
<tr>
<td>N</td>
<td>955</td>
<td>918</td>
<td>918</td>
<td>1021</td>
<td>985</td>
</tr>
<tr>
<td>Refuse/missing</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Does not know</td>
<td>109</td>
<td>109</td>
<td>109</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>Missing Cov.</td>
<td>0</td>
<td>37</td>
<td>37</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>Total Surveyed</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
<td>1081</td>
</tr>
<tr>
<td>R-squared</td>
<td>.00</td>
<td>.40</td>
<td>.49</td>
<td>.00</td>
<td>.30</td>
</tr>
<tr>
<td>Covariates?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Strata indicators?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Operator / date indicators?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**Note:** Standard errors in parentheses. ** 95% significance, * 90% significance. Dependent variables in Panel A include response to closed-ended question about the most important problem facing the country, a closed ended question about the most important problems in the Iraq war, attitudes about the leak case, the Alito confirmation, and a specific issue index constructed from the most important problem, the most important issue in Iraq and attitudes about the leak case. Dependent variables in panel B include attitudes about general national issues, including Bush approval, favorability towards Republicans, Conservatism, and a policy index constructed from these previous three items. The “broad policy index” and the “specific issue index” are both constructed by summing the standard deviations from the mean for each of the three specific questions for that index. The “Broad and Specific Issue index” is constructed then by adding together the two indices. In the row labeled “covariates”, we refer to data from the baseline survey: gender, reported age, three separate indicators for voting in the 2001, 2002 and 2004 general elections, an indicator for whether the respondent was drawn from a consumer list, self report of receiving any news or political magazines, and baseline survey self reports of preferring the Republican candidate in the gubernatorial election and having no preference in the gubernatorial election, and an indicator for wave of the study. In the row “strata indicators”, we include indicator variables for each of the strata formed prior to the randomization, which included unique combinations of the following: intention to vote, receive a paper (non-Post/non-Times), mentions ever reading a paper, gets a magazine, and asked whether they wish they read the paper more. “Surveyor/Date indicators” refers to a set of indicator variables for each unique combination of surveyor and date for the follow-up survey. All results remain qualitatively similar, and statistical significance remains as-is, using probit or ordered probit specifications instead of OLS.
Table 6: Effect of Treatment on Political Knowledge

<table>
<thead>
<tr>
<th></th>
<th>Knew number dead in Iraq</th>
<th>Identified Libby as involved in leak</th>
<th>Identified Miers as Supreme Court nominee</th>
<th>Fact Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1a) (1b) (1c)</td>
<td>(2a) (2b) (2c)</td>
<td>(3a) (3b) (3c)</td>
<td>(4a) (4b) (4c)</td>
</tr>
<tr>
<td>Post</td>
<td>-.002 (.030)</td>
<td>-.050 (.033)</td>
<td>-.057 (.031)</td>
<td>-.086 (.054)</td>
</tr>
<tr>
<td></td>
<td>.018 (.033)</td>
<td>.024 (.034)</td>
<td>.042 (.032)</td>
<td>-.034 (.054)</td>
</tr>
<tr>
<td></td>
<td>.021 (.034)</td>
<td>.022 (.034)</td>
<td>.034 (.034)</td>
<td>-.086 (.056)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.057* (.031)</td>
<td>.042 (.032)</td>
<td>-.023 (.058)</td>
</tr>
<tr>
<td>Times</td>
<td>.010 (.030)</td>
<td>.009 (.032)</td>
<td>.008 (.032)</td>
<td>-.024 (.036)</td>
</tr>
<tr>
<td></td>
<td>-.006 (.032)</td>
<td>-.011 (.034)</td>
<td>.008 (.034)</td>
<td>-.050 (.033)</td>
</tr>
<tr>
<td></td>
<td>.009 (.034)</td>
<td>.011 (.034)</td>
<td>.008 (.034)</td>
<td>-.034 (.036)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.009 (.034)</td>
<td>.008 (.034)</td>
<td>-.022 (.034)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.009 (.034)</td>
<td>.008 (.034)</td>
<td>-.050 (.034)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1077 1038 1038</td>
<td>1067 1029 1029</td>
<td>1074 1036 1036</td>
<td>1080 1041 1041</td>
</tr>
<tr>
<td>Refuse/missing</td>
<td>4 4 4</td>
<td>14 14 14</td>
<td>7 7 7</td>
<td>1 1 1</td>
</tr>
<tr>
<td>Does not know</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Missing Cov.</td>
<td>0 39 39</td>
<td>0 38 38</td>
<td>0 38 38</td>
<td>0 39 39</td>
</tr>
<tr>
<td>Total Surveyed</td>
<td>1081 1081 1081</td>
<td>1081 1081 1081</td>
<td>1081 1081 1081</td>
<td>1081 1081 1081</td>
</tr>
<tr>
<td>R-squared</td>
<td>.00 .20 .29</td>
<td>.00 .21 .32</td>
<td>.01 .23 .32</td>
<td>.00 .25 .36</td>
</tr>
<tr>
<td>Covariates</td>
<td>No Yes Yes</td>
<td>No Yes Yes</td>
<td>No Yes Yes</td>
<td>No Yes Yes</td>
</tr>
<tr>
<td>Strata indicators</td>
<td>No Yes Yes</td>
<td>No Yes Yes</td>
<td>No Yes Yes</td>
<td>No Yes Yes</td>
</tr>
<tr>
<td>Surveyor/Date indicators</td>
<td>No No Yes</td>
<td>No No Yes</td>
<td>No No Yes</td>
<td>No No Yes</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. ** 95% significance, * 90% significance. Dependent variables are: ability to identify the number dead in Iraq in a closed-ended question, identified ‘Scooter’ Libby from a list of four individuals as Dick Cheney’s chief of staff who recently resigned, identified Harriet Miers from a list of four individuals as a recent supreme Court nominee, and an index created from these questions. In the row labeled “covariates”, we refer to data from the baseline survey: gender, reported age, three separate indicators for voting in the 2001, 2002 and 2004 general elections, an indicator for whether the respondent was drawn from a consumer list, self report of receiving any news or political magazines, and baseline survey self reports of preferring the Republican candidate in the gubernatorial election and having no preference in the gubernatorial election, and an indicator for wave of the study. In the row “strata indicators”, we include indicator variables for each strata formed prior to the randomization, which included unique combinations of the following: intention to vote, receive a paper (non-Post/non-Times), mentions ever reading a paper, gets a magazine, and asked whether they wish they read the paper more. “Surveyor/Date indicators” refers to a set of indicator variables for each unique combination of surveyor and date for the follow-up survey. All results remain qualitatively similar, and statistical significance remains as-is, using probit or ordered probit specifications instead of OLS.
Appendix Table 2: Stories About the Gubernatorial Race On the Front Page or the First Metro Page

<table>
<thead>
<tr>
<th></th>
<th>Post</th>
<th>Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaine</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Kilgore</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Potts</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Kaine Ahead in Polls</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Bush Campaigns for Kilgore</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Kilgore Does not Attend Va. Bush Speech</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Kilgore Hurt by Republican party problems</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Antitax Groups Do Not Support Kilgore</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other Stories</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

* Cell entries are number of stories dealing with the gubernatorial race on each newspaper's front page from October 17, 2005 to the day of the gubernatorial election, November 8, 2005. “Kaine” refers to stories with headlines specifically about the Democratic Candidate, and “Kilgore” to stories about the Republican candidate. “Potts” refers to stories about the third party candidate.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post treatment group</td>
<td>0.003</td>
<td>-0.048</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>Times treatment group</td>
<td>0.018</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.026</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Voted in 2002</td>
<td>0.095***</td>
<td>0.103***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>From consumer database sample frame</td>
<td>0.044**</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Subscribes to news magazine</td>
<td>0.069**</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Reported preferring democratic candidate for governor</td>
<td>0.126***</td>
<td>0.126***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Wave 2 of Experiment</td>
<td>-0.037**</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Post * Female</td>
<td>0.094**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td></td>
</tr>
<tr>
<td>Post * Voted in 2002</td>
<td>-0.037</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td></td>
</tr>
<tr>
<td>Post * From consumer database sample frame</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td></td>
</tr>
<tr>
<td>Post * Subscribes to news magazine</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td></td>
</tr>
<tr>
<td>Post * Reported preferring democratic candidate for governor</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td></td>
</tr>
<tr>
<td>Post * Wave 2 of Experiment</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td></td>
</tr>
<tr>
<td>Times * Female</td>
<td>-0.040</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td></td>
</tr>
<tr>
<td>Times * Voted in 2002</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td></td>
</tr>
<tr>
<td>Times * From consumer database sample frame</td>
<td>-0.018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td></td>
</tr>
<tr>
<td>Times * Subscribes to news magazine</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td></td>
</tr>
<tr>
<td>Times * Reported preferring democratic candidate for governor</td>
<td>-0.031</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td>Times * Wave 2 of Experiment</td>
<td>-0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>3,347</td>
<td>3,347</td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.018</td>
<td>0.023</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>P(Times interaction variables $\neq 0$)</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>P(Post interaction variables $\neq 0$)</td>
<td>0.23</td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. *** 99% significance ** 95% significance * 90% significance. Indicator variable included (but not reported) if gender information is missing (applicable for 134 observations). All variables (except assignment to treatment and gender) are from the baseline survey.