**Appendix for “Increasing Precision Without Altering Treatment Effects”**

**Table of Contents**

Page 1. IRB Certificates of Approval

Page 2. Content analysis procedures

Page 3. Question wording

Page 6. Sampling details and ethics information

Page 8. Models used to estimate treatment effect size (Figure 1)

Page 14. Study 5 treatment effects by respondent partisanship

Page 16. Models used to estimate standard errors (Figure 4)

Page 22. Formal test of standard error reduction

Page 23. An alternative meta-analysis

Page 24. Details of the power analysis

**IRB Approval Certificates**

Study 1: University of Houston Committee for the Protection of Human Subjects study identification STUDY00000748

Study 2: University of Houston Committee for the Protection of Human Subjects study identification STUDY00000914

Study 3: University of Houston Committee for the Protection of Human Subjects study identification STUDY00000886

Study 4: University of Houston Committee for the Protection of Human Subjects study identification STUDY00001601

Study 5: University of Georgia Institutional Review Board study identification: PROJECT00000782

Study 8: University of Houston Committee for the Protection of Human Subjects study identification STUDY00002005

**Content Analysis Procedures**

A research assistant for one of the authors identified every article that used experiments published in the *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *Political Psychology*, and *Political Behavior* from 2015 to April of 2020. This range is inclusive of each journal’s first volume in 2020 and the first two volumes of *Political Psychology*. Articles were identified by examining the abstracts, introductions, and methods sections of each article published in the analyzed journals. Once we identified the population of articles, we analyzed a random sample of 55 publications. We discarded articles that did not include a survey experiment, leaving a sample of 41 articles. We then coded each study contained in the 41 articles, for a total of 67 studies. Each article was coded for the following attributes:

1. Use of control variables, defined as any analysis of the treatment and the outcome that includes any variable(s) that is measured but not manipulated. This could include variables like gender, education, party identification, or political values.
2. Use of a quasi measure of the dependent variable. This included variables that could be closely related to the dependent variable. For example, one study used views towards a trade policy as the outcome and included a control for support for globalization.
3. A pretreatment measure of the dependent variable. Any study that included an identical measure of the outcome measured before the participant was exposed to a treatment was coded as a pre-post design.
4. Within-subjects design. Studies were considered within-subjects if they exposed respondents to multiple experimental conditions. Notably, this includes conjoint designs, in which participants are exposed to only a small fraction of the possible conditions.

**Question Wording**

**Study 1 – Welfare Question-Wording Experiment**

*Assistance to the Poor Condition*

Generally speaking, do you think we're spending too much, too little or about the right amount on assistance to the poor? (Too much, About the right amount, Too little)

*Welfare Condition*

Generally speaking, do you think we're spending too much, too little or about the right amount on welfare? (Too much, About the right amount, Too little)

**Study 2 – Foreign Aid Information Experiment**

*Control Condition and Pretreatment Measure*

Do you think federal spending on foreign aid should be increased or decreased? (Greatly increased, Slightly increased, Kept about the same, Slightly decreased, Greatly decreased)

*Treatment Condition*

Spending on foreign aid currently makes up about 1% of the federal budget. Do you think federal spending on foreign aid should be increased or decreased? (Greatly increased, Slightly increased, Kept about the same, Slightly decreased, Greatly decreased)

**Study 3 – Education Spending Information Experiment**

*Control Condition and Pretreatment Measure*

Do you think that taxes to fund public schools should increase, decrease, or stay about the same? (Greatly increase, Somewhat increase, Stay about the same, Somewhat decrease, Greatly decrease)

*Quasi Measure*

Do you think that teacher salaries should increase, decrease, or stay about the same? (Greatly increase, Somewhat increase, Stay about the same, Somewhat decrease, Greatly decrease)

*Treatment Condition*

According to the most recent information available, $11,392 is being spent each year per child attending public schools. Do you think that taxes to fund public schools should increase, decrease, or stay about the same? (Greatly increase, Somewhat increase, Stay about the same, Somewhat decrease, Greatly decrease)

**Study 4 – Estate Tax Information Experiment**

*Control Condition and Pretreatment Measure*

Do you favor or oppose the federal estate tax on inheritances? (Strongly favor, Favor, Slightly favor, Neither favor nor oppose, Slightly oppose, Oppose, Strongly oppose)

*Quasi Measures*

Ideally, how would you prefer to see Congress attempt to reduce the federal budget deficit? (Only with spending cuts, Mostly with spending cuts, Equally with spending cuts and tax increases, Mostly with tax increases, Only with tax increases)

How strongly do you agree or disagree with the following statement?
"This country would be better off if we worried less about how equal people are."

(Strongly agree, Moderately agree, Slightly agree, Neither agree nor disagree, Slightly disagree, Moderately disagree, Strongly disagree)

*Treatment Condition*

As you may know, the federal estate tax on inheritances applies to those who have over $11.18 million, which makes up the wealthiest 0.0006% of Americans. Do you favor or oppose the federal estate tax on inheritances? (Strongly favor, Favor, Slightly favor, Neither favor nor oppose, Slightly oppose, Oppose, Strongly oppose)

**Study 5 – Prescription Drugs Party Cue Experiment**

*Control, Wave 1, and Wave 2 Pretreatment Measure*

Do you support or oppose allowing individuals to important prescription drugs from Canada? (Strongly support, Somewhat support, Slightly support, Neither support nor oppose, Slightly oppose, Somewhat oppose, Strongly oppose)

*Wave 2 Quasi Measure*

Do you support or oppose making it easier for people to import prescription drugs from other countries? (Strongly support, Somewhat support, Slightly support, Neither support nor oppose, Slightly oppose, Somewhat oppose, Strongly oppose)

*Treatment Condition*

Democrats tend to favor and Republicans tend to oppose allowing individuals to import prescription drugs from Canada. Do you support or oppose this policy? (Strongly support, Somewhat support, Slightly support, Neither support nor oppose, Slightly oppose, Somewhat oppose, Strongly oppose)

**Study 6 – GMOs Framing Experiment**

*Pretreatment Measure*

How strongly do you favor or oppose the production and consumption of genetically modified foods? (Strongly favor, Favor, Slightly favor, Neither favor nor oppose, Slightly oppose, Oppose, Strongly oppose)

*Quasi Measures*

How strongly do you favor or oppose banning the use of chemical pesticides? (Strongly favor, Favor, Slightly favor, Neither favor nor oppose, Slightly oppose, Oppose, Strongly oppose)

How strongly do you favor or oppose banning the use of antibiotics on livestock? (Strongly favor, Favor, Slightly favor, Neither favor nor oppose, Slightly oppose, Oppose, Strongly oppose)

*Anti-GMO Condition*

As you may know, opponents of genetically modified foods point out that there have not been studies on the long-term health effects of genetically modified foods on humans. And a recent study on animals found that genetically modified potatoes damaged the digestive tracts of rats.

How about you? How strongly do you favor or oppose the production and consumption of genetically modified foods? (Strongly favor, Favor, Slightly favor, Neither favor nor oppose, Slightly oppose, Oppose, Strongly oppose)

*Pro-GMO Condition*

As you may know, supporters of genetically modified foods point out that a recent study on genetically modified foods found that a type of rice (‘‘golden rice’’) can be produced with a high content of vitamin A, which is used to prevent blindness.

How about you? How strongly do you favor or oppose the production and consumption of genetically modified foods? (Strongly favor, Favor, Slightly favor, Neither favor nor oppose, Slightly oppose, Oppose, Strongly oppose)

*Recall of Attitude Change*

As you may remember, we also asked you about your support for genetically modified foods at the beginning of the survey. To the best of your memory, how has your support for genetically modified foods changed since the beginning of the survey? (Support increased since the beginning of the survey, Support decreased since the beginning of the survey, Support stayed the same)

**Sampling Details and Ethics Information**

All studies were reviewed and approved by <universities redacted for peer review> Institutional Review Boards.

*Study 1 – Welfare Question-Wording Experiment*

Respondents were recruited from mandatory introductory political science courses and from the psychology subject pool at a large public university in the southern United States. A total of 900 subjects completed the study during the spring semester of 2018. Subjects received extra credit in exchange for completing the study. Study 1 was reviewed and approved by University of Houston Committee for the Protection of Human Subjects (STUDY00000748).

*Study 2 – Foreign Aid Information Experiment*

Respondents were recruited from Amazon’s Mechanical Turk and required to be located in the U.S. and have an approval rate of at least 95%. A total of 1,209 respondents completed the study during March 25-26, 2018. Respondents were paid $0.50 for completing the survey, which is consistent with pay rates on Mechanical Turk. Study 1 was reviewed and approved by University of Houston Committee for the Protection of Human Subjects (STUDY00000914).

*Study 3 – Education Spending Information Experiment*

Respondents were recruited from Amazon’s Mechanical Turk and required to be located in the U.S. and have an approval rate of at least 95%. A total of 1,206 respondents completed the study during May 17-18, 2018. Respondents were paid $0.75 for completing the survey, which is consistent with pay rates on Mechanical Turk. Study 1 was reviewed and approved by University of Houston Committee for the Protection of Human Subjects (STUDY00000886).

*Study 4 – Estate Tax Information Experiment*

Respondents were recruited through the Lucid platform. A total of 2,462 respondents completed the survey between April 22 and May 19, 2019. Lucid aggregates respondents from many online panels and provides quota samples that are matched to US Census demographic margins on gender, ethnicity, education, region, age, and income. Respondents were compensated for their participation by the panel provider. Study 1 was reviewed and approved by University of Houston Committee for the Protection of Human Subjects (STUDY00001601).

*Study 5 – Prescription Drugs Party Cue Experiment*

Respondents were recruited through Forthright, a nationally diverse online panel maintained by Bovitz, Inc. Wave 1 (n = 1,531) was administered between July 16 and July 23, 2019. Wave 2 (n = 992) was administered between August 12 and August 19, 2019. The wave 1 sample was matched to US census characteristics on the dimensions of race, education, gender, and income. Respondents were compensated for their participation by the panel provider. Study 5 was reviewed and approved by the University of Georgia Institutional Review Board, study identification PROJECT00000782.

*Study 6 – GMOs Framing Experiment*

Respondents were recruited from mandatory introductory political science courses at a large public university in the southern United States. A total of 965 subjects completed the study during the spring semester of 2020. Subjects received extra credit in exchange for completing the study. Study 1 was reviewed and approved by University of Houston Committee for the Protection of Human Subjects (STUDY00002005).

**Models Used to Estimate Treatment Effect Size**

The six tables below display the models used to estimate the treatment effects reported in Figure 1 of the manuscript.

**Table A1. Study 1: Treatment Effects by Design**

|  |  |
| --- | --- |
|  | Posttest |
| Treatment | -0.25\*\*\* |
|  | (0.06) |
|  |  |
| Partisan ID | -0.15\*\*\* |
|  | (0.02) |
|  |  |
| Ideology | -0.01 |
|  | (0.02) |
|  |  |
| Constant | 3.01\*\*\* |
|  | (0.09) |
| *R*2 | 0.20 |
| *N* | 445 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A2. Study 2: Treatment Effects by Design**

|  |  |  |
| --- | --- | --- |
|  | Posttest | Pre-post |
| Treatment | -0.34\*\*\* | -0.13\*\* |
|  | (0.08) | (0.04) |
|  |  |  |
| Partisan ID | 0.05 | 0.04\* |
|  | (0.03) | (0.02) |
|  |  |  |
| Ideology | 0.21\*\*\* | 0.02 |
|  | (0.04) | (0.02) |
|  |  |  |
| Pretest Attitude |  | 0.87\*\*\* |
|  |  | (0.02) |
|  |  |  |
| Constant | 2.48\*\*\* | 0.23\*\* |
|  | (0.10) | (0.07) |
| *R*2 | 0.19 | 0.80 |
| *N* | 622 | 589 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A3. Study 3: Treatment Effects by Design**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Posttest | Pre-post | Quasi |
| Treatment | 0.06 | -0.03 | 0.12 |
|  | (0.09) | (0.04) | (0.09) |
|  |  |  |  |
| Ideology | -0.18\*\*\* | -0.04\* | -0.09\* |
|  | (0.04) | (0.02) | (0.04) |
|  |  |  |  |
| Partisan ID | -0.07\* | 0.01 | -0.01 |
|  | (0.04) | (0.02) | (0.04) |
|  |  |  |  |
| Pretest Attitude |  | 0.89\*\*\* |  |
|  |  | (0.02) |  |
|  |  |  |  |
| Quasi Attitude |  |  | 0.51\*\*\* |
|  |  |  | (0.06) |
|  |  |  |  |
| Constant | 4.33\*\*\* | 0.45\*\*\* | 1.82\*\*\* |
|  | (0.10) | (0.12) | (0.32) |
| *R*2 | 0.20 | 0.81 | 0.28 |
| *N* | 407 | 400 | 396 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A4. Study 4: Treatment Effects by Design**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Posttest | Pre-post | Quasi |
| Treatment | 1.03\*\*\* | 1.15\*\*\* | 1.10\*\*\* |
|  | (0.13) | (0.11) | (0.13) |
|  |  |  |  |
| Partisan ID | -0.06 | -0.05 | -0.03 |
|  | (0.03) | (0.03) | (0.04) |
|  |  |  |  |
| Ideology | -0.26\*\*\* | -0.13\*\*\* | -0.26\*\*\* |
|  | (0.04) | (0.04) | (0.04) |
|  |  |  |  |
| Pretest Attitude |  | 0.63\*\*\* |  |
|  |  | (0.03) |  |
|  |  |  |  |
| Quasi Attitude 1 |  |  | 0.28\*\*\* |
|  |  |  | (0.07) |
|  |  |  |  |
| Quasi Attitude 2 |  |  | 0.07\* |
|  |  |  | (0.03) |
|  |  |  |  |
| Constant | 4.76\*\*\* | 2.06\*\*\* | 3.78\*\*\* |
|  | (0.17) | (0.22) | (0.32) |
| *R*2 | 0.14 | 0.47 | 0.19 |
| *N* | 835 | 821 | 796 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A5. Study 5: Treatment Effects by Design**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Posttest | Pre-post | Quasi |
| Treatment | -0.44 | -0.61\*\* | -0.15 |
|  | (0.26) | (0.20) | (0.22) |
|  |  |  |  |
| Partisan ID | -0.44 | 0.01 | -0.16 |
|  | (0.29) | (0.23) | (0.23) |
|  |  |  |  |
| Treat x PID | 0.88\*\* | 0.71\*\* | 0.34 |
|  | (0.32) | (0.27) | (0.27) |
|  |  |  |  |
| Pretest Attitude (Wave 1) | 0.59\*\*\* | 0.16\*\* | 0.10\* |
|  | (0.05) | (0.05) | (0.04) |
|  |  |  |  |
| Ideology | -0.03 | 0.03 | -0.05 |
|  | (0.06) | (0.05) | (0.05) |
|  |  |  |  |
| Pretest Attitude (Wave 2) |  | 0.66\*\*\* |  |
|  |  | (0.06) |  |
|  |  |  |  |
| Quasi Attitude |  |  | 0.62\*\*\* |
|  |  |  | (0.04) |
|  |  |  |  |
| Constant | 2.59\*\*\* | 0.81 | 1.86\*\*\* |
|  | (0.51) | (0.43) | (0.43) |
| *R*2 | 0.33 | 0.51 | 0.56 |
| *N* | 289 | 287 | 296 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A6. Study 6: Treatment Effects by Design**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Posttest | Pre-post | Quasi |
| Treatment | 0.60\*\*\* | 0.60\*\*\* | 1.11\*\*\* |
|  | (0.16) | (0.12) | (0.16) |
|  |  |  |  |
| Partisan ID | 0.02 | 0.02 | -0.03 |
|  | (0.06) | (0.04) | (0.06) |
|  |  |  |  |
| Ideology | -0.02 | 0.02 | -0.01 |
|  | (0.08) | (0.06) | (0.07) |
|  |  |  |  |
| Pretest Attitude |  | 0.78\*\*\* |  |
|  |  | (0.04) |  |
|  |  |  |  |
| Quasi Attitude 1 |  |  | 0.00 |
|  |  |  | (0.07) |
|  |  |  |  |
| Quasi Attitude 2 |  |  | 0.26\*\*\* |
|  |  |  | (0.06) |
|  |  |  |  |
| Constant | 3.83\*\*\* | 0.62\*\* | 2.63\*\*\* |
|  | (0.22) | (0.23) | (0.29) |
| *R*2 | 0.04 | 0.54 | 0.20 |
| *N* | 346 | 294 | 312 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Study 5 Treatment Effects by Respondent Partisanship**

In Study 5, respondents were randomly assigned to partisan cues and we expected treatment effects to vary by respondent partisanship. Specifically, we expected Republicans would become less supportive while Democrats would become more supportive. To simplify our analysis, we presented only the interaction term in the main text. We present our finding in more detail below. First, the figure below shows the main treatment effect without moderation by respondent partisanship. As is clear, none of the main effects are statistically significant, which is unsurprising, given that the treatment effects should depend heavily on respondent partisanship. The differences between treatment effects are not statistically significant (*p*s > .09), which is particularly clear after correcting for multiple comparisons using the Holm method (*p*s > .18). Thus, we fail to reject the null of no difference in treatment effects between experimental designs.



Below, we plot the marginal treatment effects separately for Republican and Democratic respondents, for each design. As can be seen in the left panel, the treatment effect is negative for Republican respondents in all three designs, but only statistically significant in the pre-post design. However, neither the effect in the pre-post design or in the quasi design differ from the effect in the posttest design (*p*s > .34). Turning to Democrats in the right-hand panel, the treatment effect is positive and significant in the posttest design, but small and not significant in the other two designs. In this case, the treatment effect in the posttest design is larger than the effect in the pretest design (*p* = .044) and the quasi design (*p* = .041). However, because we conducted four different tests for differences between treatment effects, we correct these p-values for multiple comparisons using the Holm method. After this correction, none of the differences are statistically significant (*p*s > .16). Thus, we fail to reject the null of no differences in treatment effects between experimental designs.

****

**Models Used to Estimate Standard Errors**

The six tables below display the models used to estimate the standard errors reported in Figure 4 of the manuscript.

**Table A7. Study 1: Treatment Effects by Design**

|  |  |  |
| --- | --- | --- |
|  | Posttest | Posttest |
| Treatment | -0.120\*\*\* | -0.124\*\*\* |
|  | (0.035) | (0.032) |
|  |  |  |
| Partisan ID |  | -0.076\*\*\* |
|  |  | (0.009) |
|  |  |  |
| Ideology |  | -0.007 |
|  |  | (0.011) |
|  |  |  |
| Constant | 0.739\*\*\* | 1.007\*\*\* |
|  | (0.026) | (0.047) |
| *R*2 | 0.03 | 0.20 |
| *N* | 445 | 445 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A8. Study 2: Treatment Effects by Design**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Posttest | Posttest | Pre-post | Pre-post |
| Treatment | -0.093\*\*\* | -0.085\*\*\* | -0.033\*\* | -0.033\*\* |
|  | (0.022) | (0.020) | (0.011) | (0.011) |
|  |  |  |  |  |
| Partisan ID |  | 0.011 |  | 0.009\* |
|  |  | (0.008) |  | (0.004) |
|  |  |  |  |  |
| Ideology |  | 0.053\*\*\* |  | 0.005 |
|  |  | (0.009) |  | (0.005) |
|  |  |  |  |  |
| Pretest Attitude |  |  | 0.229\*\*\* | 0.218\*\*\* |
|  |  |  | (0.005) | (0.005) |
|  |  |  |  |  |
| Constant | 0.594\*\*\* | 0.370\*\*\* | -0.180\*\*\* | -0.192\*\*\* |
|  | (0.016) | (0.025) | (0.018) | (0.018) |
| *R*2 | 0.03 | 0.19 | 0.79 | 0.80 |
| *N* | 622 | 622 | 589 | 589 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A9. Study 3: Treatment Effects by Design**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Posttest | Posttest | Pre-post | Pre-post | Quasi | Quasi |
| Treatment | -0.01 | 0.01 | -0.01 | -0.01 | 0.04 | 0.03 |
|  | (0.03) | (0.02) | (0.01) | (0.01) | (0.02) | (0.02) |
|  |  |  |  |  |  |  |
| Partisan ID |  | -0.02\* |  | 0.00 |  | -0.00 |
|  |  | (0.01) |  | (0.00) |  | (0.01) |
|  |  |  |  |  |  |  |
| Ideology |  | -0.05\*\*\* |  | -0.01\* |  | -0.02\* |
|  |  | (0.01) |  | (0.00) |  | (0.01) |
|  |  |  |  |  |  |  |
| Pretest Attitude |  |  | 0.23\*\*\* | 0.22\*\*\* |  |  |
|  |  |  | (0.01) | (0.01) |  |  |
|  |  |  |  |  |  |  |
| Quasi Attitude |  |  |  |  | 0.15\*\*\* | 0.13\*\*\* |
|  |  |  |  |  | (0.01) | (0.02) |
|  |  |  |  |  |  |  |
| Constant | 0.63\*\*\* | 0.83\*\*\* | -0.18\*\*\* | -0.14\*\*\* | -0.01 | 0.21\* |
|  | (0.02) | (0.03) | (0.02) | (0.03) | (0.06) | (0.08) |
| *R*2 | 0.00 | 0.20 | 0.80 | 0.81 | 0.25 | 0.28 |
| *N* | 407 | 407 | 400 | 400 | 397 | 396 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A10. Study 4: Treatment Effects by Design**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Posttest | Posttest | Pre-post | Pre-post | Quasi | Quasi |
| Treatment | 0.17\*\*\* | 0.17\*\*\* | 0.20\*\*\* | 0.19\*\*\* | 0.18\*\*\* | 0.18\*\*\* |
|  | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
|  |  |  |  |  |  |  |
| Partisan ID |  | -0.01 |  | -0.01 |  | -0.01 |
|  |  | (0.01) |  | (0.00) |  | (0.01) |
|  |  |  |  |  |  |  |
| Ideology |  | -0.04\*\*\* |  | -0.02\*\*\* |  | -0.04\*\*\* |
|  |  | (0.01) |  | (0.01) |  | (0.01) |
|  |  |  |  |  |  |  |
| Pretest Attitude |  |  | 0.11\*\*\* | 0.10\*\*\* |  |  |
|  |  |  | (0.00) | (0.01) |  |  |
|  |  |  |  |  |  |  |
| Quasi Attitude 1 |  |  |  |  | 0.06\*\*\* | 0.05\*\*\* |
|  |  |  |  |  | (0.01) | (0.01) |
|  |  |  |  |  |  |  |
| Quasi Attitude 2 |  |  |  |  | 0.02\*\*\* | 0.01\* |
|  |  |  |  |  | (0.01) | (0.01) |
|  |  |  |  |  |  |  |
| Constant | 0.42\*\*\* | 0.63\*\*\* | 0.02 | 0.18\*\*\* | 0.19\*\*\* | 0.46\*\*\* |
|  | (0.02) | (0.03) | (0.02) | (0.04) | (0.04) | (0.05) |
| *R*2 | 0.06 | 0.14 | 0.45 | 0.47 | 0.13 | 0.19 |
| *N* | 837 | 835 | 821 | 821 | 796 | 796 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A11. Study 5: Treatment Effects by Design**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Posttest | Quasi | Pre-post (W1) | Pre-post (W2) |
| Treatment | -0.086 | -0.020 | -0.132\*\* | -0.087\* |
|  | (0.052) | (0.036) | (0.041) | (0.034) |
|  |  |  |  |  |
| Partisan ID | -0.014 | -0.001 | -0.003 | 0.004 |
|  | (0.048) | (0.033) | (0.040) | (0.032) |
|  |  |  |  |  |
| Treat x PID | 0.168\*\* | 0.058 | 0.106 | 0.106\* |
|  | (0.064) | (0.045) | (0.056) | (0.046) |
|  |  |  |  |  |
| Quasi Attitude |  | 0.113\*\*\* |  |  |
|  |  | (0.006) |  |  |
|  |  |  |  |  |
| Pretest Attitude (Wave 1) |  |  | 0.079\*\*\* |  |
|  |  |  | (0.009) |  |
|  |  |  |  |  |
| Pretest Attitude (Wave 2) |  |  |  | 0.125\*\*\* |
|  |  |  |  | (0.008) |
|  |  |  |  |  |
| Constant | 0.742\*\*\* | 0.129\*\* | 0.370\*\*\* | 0.039 |
|  | (0.040) | (0.041) | (0.051) | (0.051) |
| *R*2 | 0.04 | 0.54 | 0.26 | 0.49 |
| *N* | 289 | 296 | 287 | 288 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Table A12. Study 6: Treatment Effects by Design**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Posttest | Pre-post | Quasi | m4 | m5 | m6 |
| Treatment | 0.100\*\*\* | 0.099\*\*\* | 0.099\*\*\* | 0.100\*\*\* | 0.183\*\*\* | 0.185\*\*\* |
|  | (0.027) | (0.027) | (0.020) | (0.020) | (0.026) | (0.026) |
|  |  |  |  |  |  |  |
| Partisan ID |  | 0.004 |  | 0.003 |  | -0.006 |
|  |  | (0.010) |  | (0.007) |  | (0.010) |
|  |  |  |  |  |  |  |
| Ideology |  | -0.004 |  | 0.004 |  | -0.002 |
|  |  | (0.013) |  | (0.010) |  | (0.012) |
|  |  |  |  |  |  |  |
| Pretest Attitude |  |  | 0.130\*\*\* | 0.130\*\*\* |  |  |
|  |  |  | (0.007) | (0.007) |  |  |
|  |  |  |  |  |  |  |
| Quasi Attitude 1 |  |  |  |  | -0.001 | 0.001 |
|  |  |  |  |  | (0.011) | (0.011) |
|  |  |  |  |  |  |  |
| Quasi Attitude 2 |  |  |  |  | 0.043\*\*\* | 0.044\*\*\* |
|  |  |  |  |  | (0.010) | (0.010) |
|  |  |  |  |  |  |  |
| Constant | 0.470\*\*\* | 0.472\*\*\* | -0.042 | -0.063 | 0.257\*\*\* | 0.272\*\*\* |
|  | (0.019) | (0.037) | (0.031) | (0.039) | (0.042) | (0.048) |
| *R*2 | 0.04 | 0.04 | 0.54 | 0.54 | 0.20 | 0.20 |
| *N* | 346 | 346 | 294 | 294 | 312 | 312 |

Standard errors in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

**Formal Test of Standard Error Reduction**

The results presented in Figure 4 show that the pre-post designs have increased precision (smaller standard errors) compared to the post-only designs. We formally test for whether these differences are statisticallysignificant by calculating the proportion of bootstrap sample draws from the pre-post design that are smaller than the bootstrap sample draws from the post-only design. The steps of this procedure are as follows: 1) draw a sample from the post-only design and from the pre-post design, 2) record which draw is larger, and 3) calculate the proportion of pre-post draws that are *smaller* than the post-only draws. One minus this proportion is the probability that the pre-post parameter is smaller than the post-only parameter. This procedure shows that in all six studies, the probability that the pre-post standard error is larger than the post-only standard error is less than 0.0001.

**An Alternative Meta-Analysis**

We also explored an alternative approach to an internal meta-analysis that models each of the 16 treatment effects. Specifically, our dependent variable consists of the 16 treatment effects estimated from each design in each study (e.g., Study 1 provided 2 observations).[[1]](#footnote-1) For this analysis, we relied on models that did not include pretreatment measures, quasi measures, or other control variables. This choice ensured that similar weights were given to each design within each study, since estimates are weighted by precision. We then predicted these 16 observations as a function of two dummy variables for design (pre-post, quasi). Additionally, because the effect size varied across studies, we included dummy variables for each study (with Study 1 serving as the omitted case). The left panel below shows the 16 treatment effects with 95% confidence intervals (similar to Figure 1 in the manuscript). The right-hand panel shows the design coefficients from the meta-analysis. As is clear, the coefficients are both very small and statistically indistinguishable from zero.

**Figure A1. An Alternative Meta-Analytic Estimate**



**Details of Power Analysis**

In addition to the minimum detectable effect (MDE) analyses included in the main text of the paper, we also simulated a series of power analyses to illustrate consequence of the added precision in pre-post designs. The power analyses were calculated by varying up to three values used to calculate statistical power: sample size (N), the treatment effect of the experiment (measured in *Cohen’s D*), and, for the pre-post design, the correlation of the pre-treatment and post-treatment measures of the outcome. The power analysis assumes an alpha of .05 for a two-sided hypothesis test and equal sample sizes in the treatment and control groups.

The two panels in Figure 5 plot statistical power on the y-axis with sample size on the x-axis. The left panel corresponds to a standardized effect size of .2, which is roughly the smallest effect size observed in our experiments (Study 2). The red line represents estimates for the post-test design. With this effect size and design, a total sample size of roughly 780 is required to reach a common standard of 80% power. The remaining three lines depict power for the pre-post design under three different assumptions about the gain in precision from controlling for the pretreatment measure of the outcome. The green line represents the smallest gain in precision observed in our studies (a correlation between the pre-test and post-test measures of *r* = .60), the purple line represents the largest gain in precision (*r* = .90), and the blue line falls in between (*r* = .75). Even with the weakest gains, the pre-post design achieves 80% power at roughly *N* = 500, a substantial reduction from the 780 required by the posttest design. In the most optimistic scenario, the pre-post design requires only around 150 respondents to achieve 80% power. Thus, with a small effect size, which is likely common in political science, pre-post designs require a substantially smaller number of participants.

Of course, if the pre-post design also alters in the estimated coefficient, this might offset the gains in power. If common concerns are correct, and the pre-post design reduces the effect size, it will also reduce the statistical power of the test. We therefore conducted a series of simulations in which we varied the extent to which the pre-post design altered the magnitude of treatment effects. Specifically, we held all other variables constant, while shrinking the size of the coefficient in the pre-post design, relative to the posttest design. Our simulations suggest that with the weakest gains in precision from the pre-post design (*r* = .6), the pre-post design will yield the same level of statistical power as the posttest design when the pre-post design reduces the effect size to only 80% of the size of the effect in the posttest design. Under the assumption of moderate gains in precision (*r* = .75), the pre-post and posttest designs achieve equal power when the pre-post design reduces the effect size to 65%. And under the assumption of large gains (*r* = .9), the designs reach equal power when the effect size is reduced to 43% of the original effect size. Notably, these three points at which designs reach the same power are unaffected by assumptions about the original effect size or the sample size.

Overall, these results make clear the potential gains from pre-post designs. When the treatment effects remain unchanged, an assumption supported by our empirical results above, the pre-post design yields substantially greater power, requiring much smaller sample sizes. This holds true for a variety of effect sizes and for a range of empirically-based assumptions about the precision gains from pre-post designs. Finally, our results show that the pre-post design would have to change treatment effects substantially, reducing the effect size to between 80% and 43% of the original effect size, to fall to the same statistical power as the posttest design.

**Figure A2.** Simulated Statistical Power by Design-Type and Sample Size



1. For Study 1, which employed a within-subjects design, we only analyzed the second round of the experiment as a between-subjects design. This means it reaches approximately the same precision as the posttest design, which ensures it is given similar weight in the meta-analysis. [↑](#footnote-ref-1)