


RESEARCH NOTE

# Evaluating methods for examining the relative persuasiveness of policy arguments

Jared McDonald<sup>1</sup>  and Michael J. Hanmer<sup>2</sup>

<sup>1</sup>Department of Political Science and International Affairs, University of Mary Washington, Fredericksburg, VA, USA and

<sup>2</sup>Department of Government and Politics, Center for Democracy and Civic Engagement, University of Maryland, College Park, MD, USA

**Corresponding author:** Jared McDonald; Email: [jmcdona8@umw.edu](mailto:jmcdona8@umw.edu)

(Received 2 May 2023; revised 13 July 2023; accepted 22 August 2023; first published online 17 November 2023)

## Abstract

Survey researchers testing the effectiveness of arguments for or against policies traditionally employ between-subjects designs. In doing so, they lose statistical power and the ability to precisely estimate public attitudes. We explore the efficacy of an approach often used to address these limitations: the repeated measures within-subjects (RMWS) design. This study tests the competing hypotheses that (1) the RMWS will yield smaller effects due to respondents' desire to maintain consistency (the "opinion anchor" hypothesis), and (2) the RMWS will yield larger effects because the researcher provides respondents with the opportunity to update their attitudes (the "opportunity to revise" hypothesis). Using two survey experiments, we find evidence for the opportunity to revise hypothesis, and discuss the implications for future survey research.

**Keywords:** persuasion; policy attitudes; political psychology; survey experiments

## 1. Introduction

Survey experiments are an essential tool for examining reactions to political stimuli. Researchers have used them to demonstrate the public's bias against racial minority and female candidates (Terkildsen, 1993; Krupnikov and Piston, 2015; Bauer, 2020). They have shown the effect of campaign ads on attitudes toward politicians (Valentino *et al.*, 2002). And they have demonstrated that elite cues affect policy preferences (Lupia and McCubbins, 1998; Guisinger and Saunders, 2017). Survey experiments are also useful for campaigns crafting persuasive messages, activists lobbying for a policy, and pollsters seeking to understand how the public views political issues.

Yet traditional messaging experiments have limitations. In the traditional post-only design, respondents are randomly assigned to receive a single message for or against a policy prior to answering questions about that policy. These opinions are compared to those from individuals assigned to a control condition who receive no information about the policy. This approach sacrifices statistical power in order to provide "clean" estimates of public opinion. When effect sizes are small or researchers wish to investigate subgroup heterogeneity, the lack of statistical precision may be problematic.

The alternative is a repeated measures within-subjects (RMWS) design. With RMWS, respondents first provide their opinion on a policy. Next, they receive one or more arguments for or against the policy, re-answering the question after each argument. Organizations commonly use these designs when many arguments circulate in the media and among policy experts for or against a policy (e.g., Blendon *et al.*, 2003; Kirzinger *et al.*, 2021). Unlike post-only designs,

RMWS does not practically limit the number of arguments to maintain statistical power, though the potential for bias remains unexplored.

This study evaluates whether post-only and RMWS designs lead to similar conclusions in the context of policy messaging experiments. We consider two competing hypotheses. The first (opinion anchor hypothesis) suggests that the estimated effect of persuasive messages may be *smaller* in RMWS designs due to pressures respondents feel to appear consistent. The second (opportunity to revise hypothesis) suggests that the estimated effect of persuasive messages may be *larger* in a RMWS design because the researcher is explicitly offering respondents additional information and the opportunity to change their minds. We conducted experiments comparing post-only and RMWS designs using persuasive messages for two topics: (1) Washington D.C. statehood; and (2) negotiations to combat high prescription drug prices. We focus on these policy debates for two reasons. First, they represent debates in which the RMWS was previously employed, allowing us to speak to whether those findings were sensitive to the methodological approach. Second, public opinion on these issues has not calcified, thus providing the possibility for persuasion.

We find evidence that RMWS designs amplify treatment effects. Because we estimate the effect of persuasive messages, these findings are directly applicable to political practitioners (e.g., issue advocacy groups, pollsters, campaigns). Although social scientists have long worried whether measuring an outcome multiple times might lead to biased conclusions (e.g., Mutz, 2011), the present research represents one of the first attempts to empirically evaluate the magnitude of that bias (but see Clifford *et al.*, 2021).

## 2. Anchoring

Competing theories make it difficult to predict how a within-subjects approach might yield different results from a between-subjects design. On one hand, we know that individuals associate consistency with a positive self-image (Cialdini, 1984; Cialdini *et al.*, 1995). Thus, if someone answers the same question multiple times, they may be unwilling to express a change in opinion, lest expressing a different opinion reveal that the respondent's initial opinion was wrong. Consistency pressures are thought to stem from cognitive dissonance (Festinger, 1957), or the notion that individuals prefer to hold harmonious preferences. A respondent who is asked about a policy may, in that moment, develop a preference. Once confronted with an argument against the preference they just expressed, they may resist updating their beliefs. Indeed, they may reject the argument or interpret it in such a way that they do not need to arrive at a new opinion (Taber and Lodge, 2006). Stated formally:

**Opinion anchor hypothesis:** Because individuals who express an initial opinion will feel pressure to maintain consistency, when asked the same question multiple times, policy arguments will appear less persuasive in an RMWS design when compared to a post-only design. That is, respondents will tend to stick with their initial opinion when presented multiple arguments and given multiple opportunities to answer the question.

## 3. Revising

Despite consistency pressures, some features of the American electorate run counter to the anchoring hypothesis. First, most matters of policy are not central to American political decision-making (Zaller, 1992; Converse, 2006). Even if respondents express an attitude to an initial question, they may not hold that opinion deeply and might update without the discomfort of cognitive dissonance.

Clifford *et al.* (2021) discuss demand effects as a possible competing hypothesis to what we refer to as the “opinion anchor hypothesis.” According to this logic, a respondent who divines the purpose of a study may alter their response in accordance with the inferred wishes of the

researcher. Yet, as Clifford *et al.* (2021) note, the literature provides scant evidence for demand effects (Mummolo and Peterson, 2019) in survey research.

There are three reasons, however, to believe that something akin to a demand effect may exist in RMWS designs. First, influence may occur because respondents are not only likely to learn the purpose of the research, but with RMWS designs the researcher allows respondents to update their views and provides additional information to facilitate the process. Research on acquiescence bias (Krosnick and Presser, 2010) suggests that some people find disagreeing unpleasant, so they try to agree with an interviewer whenever possible. Someone seeking to be agreeable and lacking strong opinions on a policy may update their attitude after being offered an opportunity in light of new information.

Second, RMWS designs may encourage contrast effects (Schwarz and Bless, 1992). If a question is asked initially, and then re-asked in light of new information, the researcher is implicitly asking the respondent to down-weight the information they used in answering the initial question and up-weight new information. If respondents reason, “the survey would not ask for the same information twice,” they may interpret the question as asking how they feel about the argument rather than the policy.

Third, Yair and Huber (2020) and Graham and Coppock (2021) suggest that respondents engage in “response substitution.” That is, respondents answer the question they *want* to answer rather than the one that is asked. This may occur, as Gal and Rucker (2011) note, when respondents have opinions they want to convey but are not asked about. With RMWS designs, surveys provide respondents with new information but never ask respondents directly to rate that information. Instead, the questions ask respondents for their overall opinion about the policy. These theories lead to an expectation that respondents will update their reported views in an exaggerated fashion. Stated formally:

**Opportunity to revise hypothesis:** Because individuals may perceive that answering a question more than once invites them to update their views based on new information, policy arguments will appear more persuasive in RMWS designs.

These hypotheses act in opposition to one another. It is therefore possible that one will negate the effect of the other, yielding similar results. That said, the weight of the evidence from the literature leads us to favor the opportunity to revise hypothesis.

We use two studies to test these hypotheses. Each employs parallel experiments, comparing the effects of persuasive arguments using either post-only or RMWS designs. Study 1 explored DC statehood while study 2 examined prescription drug price negotiations.<sup>1</sup>

## 4. Study 1

### 4.1 Design

Study 1 was conducted from June 16–July 5, 2021, using a Qualtrics convenience sample of 1969 online panelists. The sample was designed to be representative of the US adult population with the aid of quotas for age, race, and party identification. Respondents first answered questions about their political views and news habits. In the preamble to the experiment, respondents were informed, “As you may know, Washington, D.C. is not a state and therefore does not have voting members in the U.S. Congress.” They were then randomly assigned to one of five conditions. Three conditions were associated with the traditional post-only approach while two were associated with the RMWS approach.<sup>2</sup> Respondents in conditions 1–3 took part in the post-only design, while conditions 4–5 took part in the RMWS design. The conditions were:<sup>3</sup>

<sup>1</sup>Both research designs and analyses were pre-registered ([https://osf.io/f54k8/?view\\_only=cdb846c045ea48a89326aabb449d68c3](https://osf.io/f54k8/?view_only=cdb846c045ea48a89326aabb449d68c3)). The opinion anchor and opportunity to revise hypotheses appear only in the study 2 pre-registration.

<sup>2</sup>We chose arguments based on recent news coverage of this issue.

<sup>3</sup>Complete question wording is in the Supplemental File.

- (1) *Control*. Respondents answered whether they favored or opposed DC statehood.<sup>4</sup>
- (2) *DC taxes only argument*. Respondents read that DC residents pay federal taxes before providing their opinion on DC statehood.
- (3) *DC corruption only argument*. Respondents read that DC officials have had corruption scandals before providing their opinion on DC statehood.
- (4)–(5) *RMWS*.<sup>5</sup> Respondents first answered whether they favored or opposed DC statehood. They then read the same arguments described in conditions 2 and 3, in random order, and re-answered the question about DC statehood after each.

With both studies, we examine experimental differences in two ways. First, we consider whether the effect of an argument is substantively and statistically different from the baseline in each approach. That is, we consider what conclusions we might draw about the effect of an argument if we only had the results of one research design and not both. Second, our overarching experimental assignment to the post-only or RMWS approaches allows us to evaluate the results from each design against one another. That is, we examine whether the differences in results across approaches are substantively and statistically different from one another. We recognize that substantive significance is vague and context dependent. When looking at the designs separately, we evaluate the estimated effect sizes keeping in mind whether politicians, political activists, and scholars of public opinion would care about an effect of the estimated magnitude. We also put those results into further context, evaluating whether researchers would draw the same substantive and statistical conclusions with either design. We emphasize the substantive evaluation because relying just on the statistical outcome of the difference in effect size to make a determination as to whether the differences between the two designs are important fails to recognize that a statistically insignificant difference is not the same as supporting the null hypothesis of no difference (Achen, 1982).

## 4.2 Results

In study 1, we find evidence consistent with the opportunity to revise hypothesis, i.e., the RMWS approach provides sufficient inducement for people to change their mind that overall support for the policy shifts in a meaningful way. As expected, across each approach, the baseline measure of support for DC statehood is similar (38.8 percent in the post-only versus 39.1 percent in the RMWS, see Supplementary Figure A1).<sup>6</sup> Yet upon reading the argument provided in favor of DC statehood (DC residents pay federal taxes), support goes up 5.6 percentage points ( $p = 0.11$ ) in the traditional experiment and 9.5 points ( $p < 0.01$ ) in the RMWS design (Figure 1).<sup>7</sup>

Similarly, the argument that DC sometimes elects corrupt officials dropped support for DC statehood by 2.2 percentage points ( $p = 0.53$ ) in the traditional experiment, but was 4 times larger, at 9.2 points ( $p < 0.01$ ) in the RMWS design.

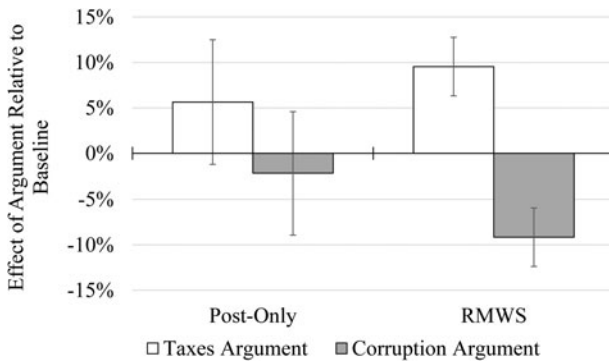
If we only had access to one of these two experiments, we would reach very different conclusions about the arguments' effects (see Figure 1). Relying on the post-only design, we would conclude that both arguments are relatively ineffective. Although the argument that DC residents pay taxes is relatively large and nearly achieves conventional levels of statistical significance, it would not be sufficient to reject the null of no effect. If we only had the RMWS design, however, we would determine that both arguments are very effective and by similar, nearly 10-point margins.

<sup>4</sup>Binary options were provided because the study was replicating previous research.

<sup>5</sup>Conditions 4 and 5 differ only in the order of the arguments.

<sup>6</sup>The similarity alleviates concern that subsequent differences are influenced by issues with the randomization.

<sup>7</sup>We calculate percent support in each condition and differences against the control/baseline within each experiment. All RMWS estimates are averaged across the first and second order. While order effects are possible, our results (across both studies) did not provide any evidence of order effects (see Supplementary Tables A3a and A3b). The same was true in a pilot study with more arguments.



**Figure 1.** Effect of arguments on support for DC statehood by research design approach.

Note: Error bars represent 95 percent confidence intervals.

Examining the difference in effect size and focusing on statistical significance, however, we find only partial evidence that the two designs yielded different results. Even though the taxes argument was nearly 4-points more persuasive in the RMWS design, that difference is not statistically significant ( $p = 0.31$ ). For the corruption argument, which had a treatment effect of 2.2 points in the post-only design compared to 9.2 points in the RMWS design, the difference is statistically significant at  $p = 0.07$ . For researchers holding to the  $p < 0.05$  threshold, that would not pass, but many would find it compelling (Achen, 1982). Nonetheless, as noted earlier, this misses several important points. First, researchers generally would choose one approach rather than run both experimentally. Here, that choice would influence the substantive conclusions they would draw. Second, analyses involving a difference in effect size require immense statistical power, since an interaction is added to estimate the difference in differences between treatment and control, both across groups and within subjects. Especially since it is reasonable to think the sample might not be large enough, focusing on substantive significance is most appropriate.

## 5. Study 2

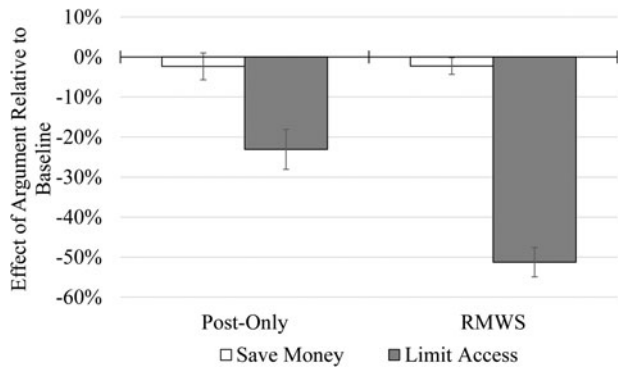
### 5.1 Design

Study 2 was conducted from November 11–12, 2021, using a convenience sample of 2002 Lucid Theorem panelists. The sample was designed to be representative of the US adult population with the aid of quotas for age, race, gender, and region. Study 2 was similar to study 1, but with a different area of policy debate—prescription drug prices. We chose a design originally used by the Kaiser Family Foundation (KFF), an organization that frequently uses the RMWS design.<sup>8</sup> This study has the advantage of having one argument that KFF found to be especially effective in shaping public opinion. The “limit access” argument described below was found to reduce support for the policy by more than 50 percentage points. As in study 1, respondents were randomly assigned to five conditions—three associated with post-only and two associated with RMWS:<sup>9</sup>

- (1) *Control*. Respondents answered whether they favored or opposed allowing the federal government to negotiate with drug companies for lower prescription drug prices, which would apply to Medicare and private insurance.
- (2) *Save money only argument*. Respondents first read that allowing the government to negotiate for lower prescription drug prices could save people money before providing their opinion on allowing such negotiations.

<sup>8</sup>KFF exclusively used an RMWS design for this issue. A full description is at: <https://files.kff.org/attachment/Topline-KFF-COVID-19-Vaccine-Monitor-May-2021.pdf>

<sup>9</sup>Complete question wording is in the Supplemental File.



**Figure 2.** Effect of argument on support for drug price negotiation by research design approach.  
Note: Error bars represent 95 percent confidence intervals.

- (3) *Limit access only argument.* Respondents first read that allowing the government to negotiate for lower prescription drug prices could limit access to new prescription drugs before providing their opinion on allowing such negotiations.
- (4)–(5) *RMWS.*<sup>10</sup> Respondents first answered whether they favored or opposed allowing the government to negotiate the price of prescription drugs. They then heard the same arguments described in conditions 2 and 3, in random order, and re-answered the item on support for allowing the negotiations.

As in study 1, we examine differences both in terms of the statistical difference and the substantive conclusions we might draw if we only had the results of one research design.

**5.2 Results**

We find that allowing the government to negotiate drug prices had exceedingly high initial support: 95.2 percent in the post-only baseline condition, and 93.5 percent in the RMWS (see Supplementary Figure A2). Support is little changed after reading the argument that such negotiations could save individuals money, likely due to a ceiling effect.<sup>11</sup>

Important differences emerge when examining the argument that such negotiations would limit access to newer prescription drugs. Although the post-only design finds that the argument is effective, reducing support for the policy by 23 percentage points to 72.1 percent, the RMWS design shows a much larger effect, similar to the effect initially discovered by KFF. Compared to the baseline, support drops all the way down to 42.2 percent—more than 51 percentage points (Figure 2).<sup>12</sup> This is consistent with the opportunity to revise hypothesis.

Again, we consider the differences between the two approaches with an eye toward both statistical and substantive significance. Here, both approaches would lead researchers to conclude

<sup>10</sup>Conditions 4 and 5 differ only in the order of the arguments.

<sup>11</sup>Although the treatment here is not strong, since the notion that prices would be lower is implicit in the control question asking about government negotiating those prices, even the most compelling argument in favor has little room for additional support given the high baseline.

<sup>12</sup>The large treatment effect provides the opportunity to explore effect heterogeneity. If the “opportunity to revise” hypothesis holds, we would expect the largest differences to exist among those whose opinions have not calcified, especially in the context of the RMWS. We chose the debate on drug price negotiation because we suspected there would be less calcification in that policy domain than others, yet it stands to reason that some people’s minds are more firmly decided than others, and thus an invitation to revise an opinion would be more influential among those with weaker opinions. Thus, we analyzed study 2 using a folded measure of ideology as a moderator. Consistent with the notion of calcification, we find that respondents who report being either “strongly conservative” or “strongly liberal” revise their opinions by less than 40 percentage points, but that less ideologically extreme respondents update their opinions by 50–60 percentage points. In the post-only experiment, we do not find similar heterogeneity. These findings are in Supplementary Table A6.



that the positive argument had little effect and the negative argument had a strong effect. Yet the RMWS approach would suggest that the limit access argument is so persuasive that, upon hearing it, a minority of Americans would support government intervention on prescription drug prices. This is far different from the 72 percent support using the traditional approach. Whether there is a solid majority (72 percent) versus a minority (42 percent) in support is substantively meaningful as it could lead to differences in how policymakers would approach the issue as well as media coverage. And whereas study 1 found only marginal statistical significance between the two approaches, the difference (23 points versus 51 points) here is also statistically significant ( $p < 0.01$ ).

## 6. Discussion

This study yielded two important findings. First, contrary to concerns that commitment and consistency pressures would make respondents unwilling to change their mind, we find many respondents are willing to update their views. That we find this in the context of both DC statehood and prescription drug prices suggests that this result is not especially sensitive to policy domain. Second, we find evidence that RMWS designs may increase the estimated effect of persuasive arguments when compared to traditional post-only designs. In sum, the findings are more consistent with the opportunity to revise hypothesis than the opinion anchor hypothesis.

Of course, our approach to the RMWS did not space out the repeated measures over the course of the survey to minimize the effect of one question on the next. Our conclusions, then, are more applicable to campaign professionals, journalists, and scholars who seek to estimate the effect of a broad range of arguments but worry that a post-only design will limit the number of arguments they can explore.

Our findings conform to expectations derived from studies of acquiescence bias (Krosnick and Presser, 2010), contrast effects (Schwarz and Bless, 1992), and response substitution (Graham and Coppock, 2021; Yair and Huber, 2020). The RMWS reduces variance but does so at the expense of accuracy. This trade-off would be unacceptable in most cases, so we advise researchers wishing to estimate accurately the effect of any individual argument to avoid this approach. Despite this, there may be some contexts where presenting respondents with multiple arguments could be appropriate. In some cases, researchers may want to provide respondents with several arguments for or against a policy simultaneously before gauging support in order to measure public opinion in an information-rich environment. This approach would require researchers to be aware of the full universe of popular arguments for and against a policy to be applicable to real-world public opinion. In other cases, researchers may want to employ the RMWS in pilot studies to identify relatively weak versus strong arguments. Future research should explore the most ecologically valid ways of examining opinions in light of multiple competing arguments.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/psrm.2023.54>. To obtain replication material for this article, <https://doi.org/10.7910/DVN/B3L9KF>

**Competing interests.** None.

## References

- Achen CH (1982) *Interpreting and Using Regression*. Washington, DC: SAGE Publications.
- Bauer NM (2020) *The Qualifications Gap: Why Women Must Be Better Than Men to Win Political Office*. New York: Cambridge University Press.
- Blendon RJ, DesRoches CM, Raleigh E and Benson JM (2003) *The Uninsured in Massachusetts: An Opportunity for Leadership*. Blue Cross Blue Shield of Massachusetts Foundation.
- Cialdini RB (1984) *The Psychology of Persuasion*. New York: Quill William Morrow.
- Cialdini RB, Trost MR and Newsom JT (1995) Preference for consistency: the development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology* 69, 318–328.
- Clifford S, Sheagley G and Piston S (2021) Increasing precision without altering treatment effects: repeated measures designs in survey experiments. *American Political Science Review* 115, 1048–1065.

- Converse PE** (2006) “The nature of belief systems in mass publics” (1964). *Critical Review* **18**, 1–74.
- Festinger L** (1957) *A Theory of Cognitive Dissonance*. Palo Alto, CA: Stanford University Press.
- Gal D and Rucker DD** (2011) Answering the unasked question: response substitution in consumer surveys. *Journal of Marketing Research* **48**, 185–195.
- Graham MH and Coppock A** (2021) Asking about attitude change. *Public Opinion Quarterly* **85**, 28–53.
- Guisinger A and Saunders EN** (2017) Mapping the boundaries of elite cues: how elites shape mass opinion across international issues. *International Studies Quarterly* **61**, 425–441.
- Kirzinger A, Kearney A, Stokes M and Brodie M** (2021) “KFF Health Tracking Poll—May 2021: Prescription Drug Prices Top Public’s Health Care Priorities.” Kaiser Family Foundation.
- Krosnick JA and Presser S** (2010) Questionnaire design. In Wright JD and Marsden PV (eds), *Handbook of Survey Research*, 2nd Edn. West Yorkshire, England: Emerald Group, pp. 263–314.
- Krupnikov Y and Piston S** (2015) Accentuating the negative: candidate race and campaign strategy. *Political Communication* **32**, 152–173.
- Lupia A and McCubbins M** (1998) *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* New York: Cambridge University Press.
- Mummolo J and Peterson E** (2019) Demand effects in survey experiments: an empirical assessment. *American Political Science Review* **113**, 517–529.
- Mutz D** (2011) *Population-based Survey Experiments*. Princeton, NJ: Princeton University Press.
- Schwarz N and Bless H** (1992) Constructing reality and its alternatives: an inclusion/exclusion model of assimilation and contrast effects in social judgment. In Martin LL and Tesser A (eds), *The Construction of Social Judgments*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 217–245.
- Taber CS and Lodge M** (2006) Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* **50**, 755–769.
- Terkildsen N** (1993) When white voters evaluate black candidates: the processing implications of candidate skin color, prejudice, and self-monitoring. *American Journal of Political Science* **37**, 1032–1053.
- Valentino NA, Hutchings VL and White IK** (2002) Cues that matter: how political ads prime racial attitudes during campaigns. *American Political Science Review* **96**, 75–90.
- Yair O and Huber GA** (2020) How robust is evidence of partisan perceptual bias in survey responses? A new approach for studying expressive responding. *Public Opinion Quarterly* **84**, 469–492.
- Zaller J** (1992) *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.