


SHORT REPORT

# Accounting for Noncompliance in Survey Experiments

Jeffrey J. Harden<sup>1</sup> , Anand E. Sokhey<sup>2\*</sup>  and Katherine L. Runge<sup>2</sup>

<sup>1</sup>Department of Political Science, University of Notre Dame, 2055 Jenkins Nanovic Halls, Notre Dame, IN 46556, USA, e-mail: [jeff.harden@nd.edu](mailto:jeff.harden@nd.edu) and <sup>2</sup>Department of Political Science, University of Colorado Boulder, 333 UCB, Boulder, CO 80309, USA, e-mail: [katherine.runge@colorado.edu](mailto:katherine.runge@colorado.edu); Twitter: @AESokhey  
\*Corresponding author. Email: [anand.sokhey@colorado.edu](mailto:anand.sokhey@colorado.edu)

## Abstract

Political scientists commonly use survey experiments—often conducted online—to study the attitudes of the mass public. In these experiments, compensation is usually small and researcher control is limited, which introduces the potential for low respondent effort and attention. This lack of engagement may result in noncompliance with experimental protocols, threatening causal inferences. However, in reviewing the literature, we find that despite the discipline’s general familiarity with experimental noncompliance, researchers rarely consider it when analyzing *survey* experiments. This oversight is important because it may unknowingly prevent researchers from estimating their causal quantities of greatest substantive interest. We urge scholars to address this particular manifestation of an otherwise familiar problem and suggest two strategies for formally measuring noncompliance in survey experiments: recording vignette screen time latency and repurposing manipulation checks. We demonstrate and discuss the substantive consequences of these recommendations by revisiting several published survey experiments.

**Keywords:** Survey experiments; noncompliance; causal estimands

Researchers who conduct survey experiments often encounter respondents who do not pay sufficient attention to questions and experimental vignettes. This problem is especially acute in the increasingly common online setting (Berinsky, Margolis, and Sances 2014). Experimental noncompliance is not a new problem, and methods for addressing it are well known. However, we find that current practices for analyzing survey experiments rarely include these methods. Accordingly, we encourage the use of two simple strategies for identifying noncompliant survey respondents. These strategies can be used with established methods for estimating treatment effects under noncompliance.

---

The data, code, and any additional materials required to replicate all analyses in this article are available at the Journal of Experimental Political Science Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/MNK26U> (Harden, Sokhey, and Runge 2018). The authors have no conflicts of interest.

© The Experimental Research Section of the American Political Science Association 2019.

Table 1  
 Noncompliance in Political Science Survey Experiments, 2006–2016

Journal	Articles	Mention noncompliance	Mention estimand change	Report multiple estimands
<i>American Political Science Review</i>	11	2 (18%)	1 (9%)	0 (0%)
<i>American Journal of Political Science</i>	32	6 (19%)	5 (16%)	3 (9%)
<i>Journal of Politics</i>	25	1 (4%)	1 (4%)	1 (4%)
<i>Public Opinion Quarterly</i>	34	6 (18%)	1 (3%)	1 (3%)
<i>Political Behavior</i>	28	2 (7%)	0 (0%)	0 (0%)
Total	130	17 (13%)	8 (6%)	5 (4%)

Note: Cell entries summarize our coding of articles reporting survey experiments published from 2006 to 2016. See the Appendix in the Supplementary Material for complete details of the coding procedure.

## The State of the Literature

We conducted a meta-analysis of all articles reporting survey experiments published from 2006–2016 in five major political science journals. We coded whether each article mentions noncompliance, discusses the fact that noncompliance changes the causal estimand, or reports multiple estimands due to noncompliance. We summarize the results in Table 1 (see the Appendix in the Supplementary Material for further details).

The results indicate that most articles published in these major journals ignore the possibility of noncompliance in survey experiments. Additionally, some of those that discuss the issue respond to it by subsetting the data or dropping noncompliant respondents. These ad-hoc practices are problematic because they usually amount to conditioning on a post-treatment variable (see the Appendix in the Supplementary Material). In those cases, subsetting will alter the causal estimand and introduce bias into the estimate.

The common approach of ignoring noncompliance typically yields an estimate of the intent-to-treat effect (ITT). However, in practice, some researchers who report the ITT seem to believe that they have identified the average treatment effect (ATE). This practice can be problematic if noncompliance is present; estimating the ITT but interpreting it as the ATE is inconsistent with the goal of conceptual precision in causal analysis. Of course, a researcher conducting a survey experiment may actually be interested in treatment effectiveness, which is the substantive interpretation of the ITT. However, we contend that in such a case, the researcher should provide justification for the choice to focus on the ITT. Additionally, even if estimates of the two quantities are signed similarly, the researcher should care about accurately estimating the magnitude of a treatment effect.

## Recommendations

We propose two straightforward methods for measuring noncompliance in survey experiments. These are somewhat simple measures and are certainly not the only

possibilities. However, they are useful because they are easy to implement and effective at separating compliers from noncompliers. After employing one or both of them to measure noncompliance, a typical approach is to estimate the causal effect among compliers. Researchers often employ instrumental variables analysis to estimate the complier average causal effect (CACE) in such a case (see Angrist, Imbens, and Rubin 1996).

Our first suggested method for measuring compliance is to record screen time for the display of experimental vignettes. As with other “latency measures,” the analyst can determine some minimum acceptable time at which to code a respondent as compliant. Prior research on reading speed may help with this decision (see Rayner 1998). However, as a best practice, we recommend the use of pilot studies on small convenience samples to learn more about the time it takes for respondents to read and internalize vignettes in the specific experiment in question. In the Appendix, we provide a detailed demonstration of this approach with a replication of a survey experiment in Harden (2016).

Second, researchers can repurpose manipulation checks to determine whether a respondent read and thought about the treatment. A manipulation check involves one question or a series of questions designed to evaluate whether respondents can display comprehension of the vignette content. Such questions are typically used to assess internal validity of the experimental manipulation, but they can also be designed to measure compliance. We also recommend the use of pilot studies with this approach. Testing manipulation checks can help researchers with decisions such as whether to use a single question or a battery of several questions as well as the number of correct answers a respondent must provide to qualify as a complier.

Both of these strategies require the researcher to make some seemingly arbitrary decisions, such as the latency cut-off time or the number of correct answers in a manipulation check. While pilot testing can help with this issue, it likely will not remove it completely. Accordingly, we also recommend that researchers describe and justify their chosen strategy for measuring compliance in pre-analysis plans (see Monogan 2013). Publicly committing to a particular strategy before collecting data holds the researcher accountable and minimizes the risk of adjusting the definition of compliance after looking at results.

## The Bottom Line

Noncompliance is likely present in many survey experiments, particularly given the popularity of recruiting respondents from online pools. This is an old problem in a new form, but one that poses real threats to securing causal inferences and drawing meaningful substantive conclusions. In the Appendix, we present replications of 51 treatment effect estimates from survey experiments in six published studies. We find that moving from the ITT to the CACE corresponds with a median increase of 28% in the magnitude of the effects. In sum, we urge scholars to pay attention to indicators of noncompliance in this familiar setting and to properly account for it when the problem is present.

**Supplementary materials.** To view supplementary material for this article, please visit <https://doi.org/10.1017/XPS.2019.13>

Author ORCID*s*.  Anand E. Sokhey <https://orcid.org/0000-0002-3403-0967>, Jeffrey J. Harden <http://orcid.org/0000-0001-5337-7918>

## References

- Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin.** 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91(434):444–55.
- Berinsky, Adam J., Michele F. Margolis and Michael W. Sances.** 2014. Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys. *American Journal of Political Science* 58(3):739–53.
- Harden, Jeffrey J.** 2016. *Multidimensional Democracy: A Supply and Demand Theory of Representation in American Legislatures*. New York: Cambridge University Press.
- Harden, Jeffrey J., Anand E. Sokhey and Katherine L. Runge.** 2018. Replication Data for: Accounting for Noncompliance in Survey Experiments. *Harvard Dataverse Network*, V1. doi: [10.7910/DVN/MNK26U](https://doi.org/10.7910/DVN/MNK26U).
- Monogan, James E.** 2013. A Case for Registering Studies of Political Outcomes: An Application in the 2010 House Elections. *Political Analysis* 21(1):21–37.
- Rayner, Keith.** 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin* 124(3):372–422.

---

**Cite this article:** Harden JJ, Sokhey AE, and Runge KL (2019). Accounting for Noncompliance in Survey Experiments. *Journal of Experimental Political Science* 6, 199–202. <https://doi.org/10.1017/XPS.2019.13>